Diagnosis of Regime-dependent Cloud Simulation Errors in CMIP5 Models Using "A-Train" Satellite Observations and Reanalysis Data

Hui Su, Jonathan H Jiang, Chengxing Zhai, Vince S. Perun, Janice T. Shen
Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, California, USA

Anthony Del Genio and Larissa S. Nazarenko
Goddard Institute for Space Studies (GISS), New York, New York, USA

Leo J. Donner, Larry Horowitz, Charles Seman
Geophysical Fluid Dynamics Laboratory (GFDL), Princeton, New Jersey, USA

Cyril Morcrette, Jon Petch, Mark Ringer
UK Met Office (UKMO) Hadley Center, London, UK

Jason Cole
Canadian Centre for Climate Modeling and Analysis (CCCMA), Environment Canada, Toronto, Canada

Michel d. S. Mesquita
Bjerknes Centre for Climate Research (BCCR), Uni Research, Bergen, Norway

Trond Iversen
Norwegian Climate Centre (NCC), Meteorologisk Institutt, Oslo, Norway

Jon Egill Kristjansson
University of Oslo, Oslo, Norway

Andrew Gettelman
National Center for Atmospheric Research (NCAR), Boulder, Colorado

Leon Rotstayn
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Aspendale, Australia

Stephen Jeffrey
Queensland Climate Change Centre of Excellence (QCCCE), Queensland, Australia

Jean-Louis Dufresne
Laboratory of Dynamical Meteorology, Institute Pierre Simon Laplace (IPSL), France

Masahiro Watanabe
Model for Interdisciplinary Research On Climate (MIROC)
Atmospheric and Ocean Research Institute, University of Tokyo, Chiba, Japan

Hideaki Kawai and Tsuyoshi Koshiro
Meteorological Research Institute (MRI), Japan Meteorological Agency, Tsukuba, Japan

Tongwen Wu
Beijing Climate Center (BCC), China Meteorological Administration, Beijing, China

Evgeny M. Volodin
Institute for Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia

Tristan L'Ecuyer
University of Wisconsin-Madison, Madison, Wisconsin, USA

Joao Teixeira, Graeme L. Stephens
JPL, California Institute of Technology, Pasadena, California, USA

Key Words: Clouds, Climate Model, and Satellite Observation;

Abstract. The vertical distribution of cloud water content (CWC) and cloud fraction (CF) over the tropical oceans, produced by 13 coupled atmosphere-ocean models submitted to the Phase 5 of Coupled Model Intercomparison Project (CMIP5) are evaluated against CloudSat/CALIPSO observations as a function of large-scale parameters. Available CALIPSO simulator CF outputs are also examined. A diagnostic framework is developed to decompose the cloud simulation errors into the large-scale errors, cloud parameterization errors and co-variation errors. We find that the cloud parameterization errors contribute predominantly to the total errors for all models. The errors associated with large-scale temperature and moisture structures are relatively greater than those associated with large-scale mid-tropospheric vertical velocity and lower-level divergence. All models capture the separation of deep and shallow clouds in distinct large-scale regimes; however, the vertical structures of high/low clouds and their variations with large-scale parameters differ significantly from the observations. The deep convective clouds simulated in most models do not reach as high in altitude as observed, and their CWC are generally weaker in magnitude than CloudSat total CWC, which includes the contribution of precipitating condensates, but are close to CloudSat non-precipitating CWC. All models reproduce maximum CF associated with convective detrainment, but CALIPSO simulator CFs agree better with CloudSat/CALIPSO combined retrieval than the model CFs, especially in the middle troposphere. Model simulated low clouds tend to have little variation with large-scale parameters except for lower-troposphere stability, while the observed low cloud CWC, CF and cloud top height vary consistently in all large-scale regimes.

1. Introduction

The simulations of clouds by the nearly 20 models submitted to the Phase 5 of Coupled Model Intercomparison Project (CMIP5) exhibit large discrepancies from satellite observations [*Jiang et al*., 2012]. Despite a number of models displaying a certain degree of improvement in some aspects of the cloud simulations (e.g., the column-integrated cloud water path), no substantial improvements across the models are found compared to those submitted to the Phase 3 of CMIP (CMIP3), which constitute the basis for the climate projections in the Fourth Assessment Report (AR4) by the Intergovernmental Panel for Climate Change (IPCC) [*Randall et al*., 2007]. Hence, improving the representation of clouds in climate models still poses a great challenge. While the spatial resolution of climate models will likely increase substantially in the years ahead (e.g., from hundreds of kilometers to tens of kilometers in horizontal resolution), many cloud processes will still be parameterized because they occur on sub-grid scales. To help validate and formulate cloud parameterizations, it is important to examine the relationships between observed cloud properties and their environment and then compare these observed relationships with those simulated by models. This study documents the relationship between the cloud vertical structures and large-scale parameters in CMIP5 model simulations, and evaluates these relationships against those from state-of-the-art satellite observations. It complements our earlier work [*Jiang et al*., 2012], in which the CMIP5 model performance in simulating water vapor and clouds over the tropical oceans was quantified in terms of spatial mean, variance and correlation. One particular question we address is how much discrepancy of simulated clouds from observations is caused by errors in simulating large-scale parameters and how much is caused by cloud parameterizations used by the models.

Two cloud properties are examined in our analysis: the vertically resolved cloud water content (CWC) and cloud fraction (CF), both being important variables in determining cloud radiative effect (CRE), although their treatments in radiation codes may vary among models. Errors stemming from either variable have varying degree of impact on CRE calculations. Thus, it is necessary to inspect both quantities. To our knowledge, there have been parallel efforts in evaluating CMIP5 simulated clouds in terms of cloud fraction, cloud optical depth, cloud top height, cloud occurrence frequency and simulated satellite radiances (such as reflectivity) (e.g., S. Klein and group, *personal communication*, 2012). The multi-facets of cloud simulations revealed by different cloud variables will provide comprehensive knowledge of cloud processes and help to accelerate the progress of improving cloud simulations in models, contributing to the reduction of uncertainties in climate change predications.

This model diagnosis work primarily utilizes the "conditional sampling approach" originated by *Bony et al.* [2004], in which cloud variables are sorted into bins of large-scale parameters, representing distinct large-scale dynamic and thermodynamic regimes. These large-scale variables may directly or indirectly enter the cloud parameterization schemes to govern cloud formation, growth and dissipation, and they also vary simultaneously as clouds evolve. A number of studies have applied this approach to address cloud and water vapor variabilities and feedbacks [*Bony and Dufresne*, 2005; *Huang et al.*; 2006; *Williams et al*., 2003; *Wyant et al*., 2006; *Bennhold and Sherwood*, 2008; *Zhang et al*., 2007; *Su et al.*, 2008; *Su et al.*; 2011]. In *Su et al.* [2008], eight large-scale quantities were examined in relation to CloudSat observed CWC profiles. Clear clustering of cloud structures in large-scale regimes was revealed. In this study, we use six large-scale environmental parameters, including mid-tropospheric vertical velocity at 500 hPa (ω_{500}), lower-level divergence (LDV), sea surface temperature (SST), lower tropospheric stability (LTS),

water vapor path (WVP), and relative humidity (RH) to sort CWC and CF profiles. We also conditionally sample clouds by precipitation so as to evaluate how models capture the co-variability of the two products out of moist processes. In this study, we do not use SST gradient and convective available potential energy (CAPE) as in *Su et al.* [2008] because the cloud profiles sorted by these two variables are not as distinctly clustered as by other variables [*Su et al.*, 2008]. Using a number of large-scale parameters offers multiple perspectives of the cloud processes and demonstrates clear benefits over using a single parameter, as we will show later. A "good" model performance in one large-scale regime does not guarantee a good performance in another regime.

When comparing the cloud distributions in large-scale regimes between observations and model simulations, one may wonder whether and how much the departures from observations are caused by the imperfect simulations of large-scale parameters. *Su et al.* [2011] used three large-scale variables, ω_{500}, SST and LTS, to sort cloud simulations in two climate models and one re-analysis system. They showed that the clouds produced by the re-analyses system match closer to the CloudSat observation than the two free-running models, demonstrating that more stringent constraint on grid-scale state variables helps to produce better cloud fields. For CMIP5 atmosphere-ocean coupled models, the large-scale environmental variables are governed by internal model dynamics and could have sizeable deviations from observations. Such large-scale errors may propagate into cloud errors through cloud parameterizations. On the other hand, our insufficient knowledge of cloud processes translates into unjustified *ad hoc* assumptions in cloud parameterizations, causing significant errors in simulated clouds even though the large-scale states are correctly simulated. A primary goal of this study is to examine the relative contributions of large-scale errors and cloud parameterization errors. To achieve this, we devise an innovative

diagnostic framework to decompose the two types of errors and quantify their relative importance. Such decomposition is particularly useful to identify target areas for future model improvements.

The structure of the paper is as follows. Section 2 describes the observational datasets, models and the diagnostic approach. Section 3 presents the comparison of the regime-sorted cloud profiles and the decomposition of cloud errors. Concluding remarks are given in Section 4.

2. Observations, Models and Diagnostic Approach

2.1 Observations

We use CWC profiles from CloudSat 2B-CWC-RO (V4) and CF from combined radar and lidar retrievals, version 4 [*Austin et al.*, 2009]. The uncertainties for CWC and CF are a factor 2 and ~5%, respectively [*Jiang et al.*, 2012]. For CloudSat CWC retrieval, the inclusion of precipitating particles is a major concern when comparing with model outputs that may or may not include rain or snow in the reported ice water content (IWC) or liquid water content (LWC) [*Jiang et al.*, 2012; *Waliser et al.*, 2009]. *Su et al.* [2011] showed that the vertical structure of non-precipitating CWC sorted by various large-scale variables does not differ significantly from total CWC. Thus, we use the total CWC sorted by large-scale parameters to compare with the modeled counterparts for discussions of cloud vertical structure. For a quantitative estimate of model biases, we use non-precipitating and total CWC as the lower and upper bounds of observed CWC, respectively. To conduct the conditional sampling, we construct monthly CWC and CF at horizontal grids of 2.5º (longitude) × 2º (latitude) each and onto 40 vertical levels from the surface to 20 km, with an interval of 500 m. The monthly CloudSat/CALIPSO data from August 2006 to July 2010 are used. As CloudSat/CALIPSO observations are made twice a day (at ~1:30am and ~1:30pm local time), we limit our analysis to tropical (30°S-30°N) oceanic regions where the

impact of cloud diurnal cycle on the monthly mean data is relatively small. The quality of satellite retrievals is generally better over ocean than over land [*Austin et al*., 2009].

The observed mid-tropospheric vertical velocity at 500 hPa (ω_{500}) is taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis to represent large-scale circulation. Low-level divergence (LDV) is defined as the mean horizontal divergence within 1000 hPa and 850 hPa (i.e., $LDV = \frac{1}{dP}\int_{850}^{1000}(\nabla \cdot \vec{V})dp$, where dP=150 hPa and $\vec{V}$ is the horizontal winds), calculated from ECMWF interim reanalysis. LTS is calculated as the potential temperature difference between 1000 and 700 hPa (LTS = $\theta_{1000} - \theta_{700}$) based on the temperature measurements from the Atmospheric Infrared Sounder (AIRS) on Aqua. For SST and WVP, we use the retrievals from the Advanced Microwave Scanning Radiometer (AMSR-E) on Aqua, downloaded from the Remote Sensing Systems (http://www.remss.com) with a horizontal resolution of 0.25°×0.25°. We also use the relative humidity with respect to water (ice for pressures less than 300 hPa) from ECMWF interim reanalysis to sort cloud profiles. We explored using RH data from AIRS to sort cloud profiles but found that the low bias of AIRS RH data in heavily cloudy regions limits its use. The observed precipitation data are taken from the Global Precipitation Climatology Project (GPCP) monthly product from January 1980 to December 2004. The temporal coverage for satellite data depends on respective mission periods. For conditional sampling, we re-grid the original observed large-scale datasets onto horizontal grid boxes of 2.5º (longitude) × 2º (latitude), same as the CWC and CF. Details about the satellite data used can be found in *Jiang et al*. [2012].

2.2 CMIP5 models

While 19 models were examined in *Jiang et al.* [2012], only 13 of them are used in this study for two reasons. First, we select only models that, at least, are atmosphere-ocean coupled, as they are used for future climate predictions. NCAR cesm1-cam5 is an atmosphere-ocean coupled model but

is not yet included because many of its large-scale variables are not archived on the ESG at the time of this study. The atmosphere-only model simulations use prescribed SST, and thus are not fair comparisons with coupled model simulations. Second, we choose models that have archived both CWC and CF at the Earth System Grid (ESG) hosted by the Program for Climate Model Diagnosis and Intercomparison (PCMDI). All model outputs are from the last 25 years of "the historical runs" [*Taylor et al*., 2012] from 1980-2004. Original model outputs are re-gridded to the same 2.5º×2º horizontal grid boxes and 40 vertical levels as done for the observed clouds. CSIRO mk3.6 model outputs of large-scale state variables are limited to 920 hPa and above. Thus, its LDV refers to the averaged divergence between 920 and 850 hPa. To calculate its LTS, we use SST instead of air temperature at 1000 hPa, which is used for all other models. The model descriptions and key references can be found in Table 1 of *Jiang et al.* [2012].

In some of the models, satellite simulators are implemented. The Cloud Feedback Model Intercomparison Project (CFMIP) has developed an integrated satellite simulator, the CFMIP Observation Simulator Package (COSP) for use in many CMIP5 models [*Bodas-Salcedo et al.,* 2011]. These simulators take into account instrument sensitivities and can circumvent retrieval errors when their outputs of satellite radiances are directly compared with satellite measurements (e.g., radar reflectivity, brightness temperature). *Bodas-Salcedo et al.* [2008] showed useful application of CloudSat simulator in evaluating model simulated cloud structures. However, differences in the simulated radiances from the simulators in models compared to the satellite measurements may not tell the differences in the modeled CWC from the "true" CWC because the "true" cloud particle size and shape are unknown and assumptions for them in the simulators are model-dependent. It is possible to obtain very similar radiances from the simulators given two different CWC profiles by using different cloud particle size assumptions. For CWC, comparing

model outputs with the CloudSat retrieval is probably not more uncertain than comparing reflectivity and then inferring CWC [*Steve Klein* 2012, personal communication]. For CF, however, CloudSat/CALIPSO simulator CF may be a more fair choice for comparison with the radar-lidar combined CF as instrument sensitivity is one of the most important factors in cloud detection which determines CF, other than cloud particle size and shape assumptions. At the time of analysis, only six models (CCCMA am4, CNRM cm5, IPSL cm5a, MIROC miroc5, MRI cgcm3, and UKMO hadgem2-es) on the ESG provide CALIPSO simulator CF and two models have CloudSat simulator CF, while many provide the International Satellite Cloud Climatology Project (ISCCP) simulator CF. For purposes of illustration, we examine both the original and CALIPSO simulator CF for the six models, and use only the original CF for the rest of models. We plan to add more simulator CF results when the outputs become available on the ESG.

Figure 1 shows the zonal-mean and tropical-mean (30°N-30°S) original and CALIPSO simulator CF for the six models. For the zonal means, it is clear that all simulator CFs are less than the original model CFs in most of the free troposphere and at all latitudinal bands, except that CCCMA am4 simulator CF shifts higher in altitude than the original CF in the tropical upper troposphere. The lower values of the CALIPSO simulator CF than the original CF likely result from the inability of CALIPSO to see through thick clouds. For the tropical means, the CALIPSO simulator CF for CCCMA am4 is greater than the original CF at pressure levels above 200 hPa and in the boundary layer, but is significantly smaller than the original CF between 800 hPa and 200 hPa. For all other models, the simulator CF is smaller than the original CF over most of the troposphere, except that IPSL and MIROC models have the simulator CF about 5% greater than the original CF at 900 hPa. For MRI cgcm3 and UKMO hadgem2-es, the CALIPSO simulator CF and the original CF are nearly identical at altitudes higher than 200 hPa. Apparently, instrument

sensitivity matters in the quantitative comparison of CF. An ideal choice for the comparison against the radar-lidar combined CF retrieval should be combined radar-lidar simulator CF, which is not available at present. Hence, using CALIPSO simulator CF alone does not fully address the "fairness" of model-data comparison. Nevertheless, the differences between the simulator CF and original CF shown in Figure 1 illustrate the uncertainties associated with satellite retrievals and caveats using original CF for quantitative comparison with observations. We shall bear in mind these caveats when interpreting the model-observation discrepancies.

2.3 Diagnostic framework

The essence of the conditional sampling approach is to treat tropical clouds as a function of large-scale regime parameters, in lieu of geographical spatial coordinates. We denote the observed and model simulated cloudiness (CWC or CF) in a large-scale regime of the variable V as C_v^o and C_v^m, respectively, and the occurrence frequency (probability density function, pdf) of the large-scale regime V as P_v^o for observations and P_v^m for models, with $\int_{-\infty}^{+\infty} P_v^{o,m} dV = 1$. The tropical-averaged cloudiness, <C>, can be expressed as an integral of cloudiness in each regime over the entire range of the V values, i.e.,

$$< C^{o,m} >= \int_{-\infty}^{+\infty} P_v^{o,m} C_v^{o,m} dV . \quad (1)$$

The difference between the model-simulated and observed cloudiness in each regime V is $\delta(P_v C_v) = P_v{}^m C_v{}^m - P_v{}^o C_v{}^o$, which can be decomposed into three components,

$$\delta(P_v C_v) = C_v^o \cdot \delta P_v + P_v^o \cdot \delta C_v + \delta P_v \cdot \delta C_v \quad . \quad (2)$$

In eq. (2), $\delta P_v = P_v^m - P_v^o$ is the difference between the modeled and observed pdf of large-scale regime V and the first term on the right hand side (r.h.s.) represents the cloud errors

associated with inaccurate simulations of large-scale parameters assuming accurate cloud parameterizations (C_v^o) – we call it the "large-scale error". The second term on the r.h.s. of eq. (2) represents the cloud errors due to the incorrect cloud parameterizations ($\delta C_v = C_v^m - C_v^o$) given the correct large-scale parameter pdf (P_v^o) – we call it the "cloud parameterization error". The last term is the co-variation of the first two error sources. Integrated over the possible values of V on the tropics, we can compute the total errors on the tropical mean and the contributions of each component to the total errors.

3. Results

3.1 Sorting clouds by dynamical variables ω_{500} and LDV

3.1.1 Correlation between ω_{500} and LDV

Although it is beneficial to examine the cloud structure in many different large-scale parameters, it is well-known that these large-scale parameters are not independent. Here, we examine to what extent they are correlated so that we have an idea about how much additional information may be provided by using multiple parameters for conditional sampling. For dynamical regimes, we choose ω_{500} and LDV as two regime indicators. The scatter plot in Figure 2 demonstrates they are indeed correlated with a correlation coefficient of 0.77. The upward motion corresponds to lower-level convergence and vice versa. Noticeable scatter exists, especially over intermediate circulation regions (–40 hPa/day < ω_{500} < 20 hPa/day), where a wide range of LDV values (either divergence or convergence) are associated with the same vertical velocity. As 40% of the variance of LDV is not explained by $\omega_{500,}$ examining clouds in the functional space of LDV would provide some new information in addition to sorting clouds by ω_{500}.

3.1.2 Cloud profiles sorted by ω_{500}

Since *Jiang et al.* [2012] evaluated the model performance in simulating the magnitudes of CWC and CF at discrete pressure levels, our attention in this paper is focused on the vertical variations of CWC and CF in large-scale parameter regimes. Figure 3 displays the modeled CWC and CF (as well as simulator CF, if available) as a function ω_{500} for CloudSat/CALIPSO observation and 13 CMIP5 models. The horizontal map of the tropical ω_{500} from ECMWF averaged for the CloudSat/CALIPSO observation period (2006-2010) is shown on the lower-right corner, and the differences between the model simulated and the ECMWF reanalysis ω_{500} (25-year average) are shown under each model. The observed CWC and CF are grouped into two clusters: one associated with deep convective clouds in the ascending regimes and one associated with stratiform clouds in the descending regimes. As ω_{500} changes from positive to negative, CWC and CF amounts in the boundary layer (below 800 hPa) decrease, while their amounts in the middle and upper troposphere (UT) increase. The height of maximum CWC is lower than the height of maximum CF because of the high sensitivity of CALIPSO lidar to thin cirrus. All models capture the two dominant modes of clouds (high and low clouds); however, the variations in the vertical and with ω_{500} differ substantially from the observations.

In the large-scale ascending regimes ($\omega_{500} < 0$), out of the 13 models, only two GISS models (e2-h and e2-r, the same atmospheric model coupled to two different ocean models) reproduce the high CWC in the UT around 200 hPa, although its magnitude is higher than the observed total CWC. The CF for high clouds from GISS is consistent with CWC, with its peak around 200 hPa, lower than the lidar-radar combined retrieval. The CALIPSO simulator CF for GISS is not available, which might be different from the original CF. GFDL cm3 deep convective clouds spread a broad layer from the middle troposphere to the UT; however, its maximum CWC peaks at 500 hPa, lower in altitude than CloudSat CWC. All other models do not produce CWC peaking

in the UT; instead, their CWCs maximize between 600 and 400 hPa. CCCMA canesm2 and CNRM cm5 have their maximum CWCs lower than 800 hPa over the ascending regimes. CSIRO mk3.6 and UKMO hadgem2-es have comparable CWC magnitudes for deep clouds, both being quite weak and reaching lower in altitude than the observed. On the other hand, all models produce maximum CF around 200 hPa, where convective detrainment prevails. Most models tend to have a secondary maximum CF in the lower troposphere between 700 and 800 hPa, while INM cm4 and MIROC micro5 have a distinct peak CF around 600 hPa in addition to the maximum CF around 200 hPa. The INM cm4 has the weakest CWC in the large-scale ascending regimes among all models. Its low bias in CWC is compensated by its high bias in CF. We speculate that the modeled CF may have been heavily adjusted in some models because of its direct relevance to CRE calculations. For CWC, less attention may have been paid and there is more freedom in the models as some radiation codes do not explicitly take into account the values of CWC. For the models that produce CALIPSO simulator CF, the simulator CFs are noticeably closer to the radar-lidar combined retrieval in the middle troposphere and UT, confirming the usefulness of applying simulator for CF comparison. The MIROC miroc5 has a high bias of CF in the middle troposphere, which is not alleviated by using its CALIPSO simulator CF. For models that do not have CALIPSO simulator CF, there appear to be high biases in CF from the middle troposphere to 200 hPa, but low biases in CF above 200 hPa.

In the descending regimes ($\omega_{500} > 0$), all models rightly produce clouds in the lower troposphere, although the CWC and CF magnitudes, the cloud top heights and how CWC and CF change with ω_{500} vary significantly between the models and the observation. INM has the lowest low cloud CWC among all models. MRI cgcm3 also has little stratiform clouds, with a local maximum around ω_{500} of 20-30 hPa/day, unlike the observed low cloud CWC maximizing over

ω_{500} > 50 hPa/day. Most models do not reproduce the decrease of CWC in the boundary layer when the subsidence rate decreases, except CSIRO mk3.6 and GFDL cm3, despite that its magnitudes in CWC and CF are different from the observations. All modeled CWCs in the descending regimes are smaller than CloudSat total CWC but greater than CloudSat non-precipitating CWC (not shown). The simulated low cloud CFs are generally smaller than the observed, except for INM cm4 and MIROC miroc5. The MIROC miroc5 CALIPSO simulator CF is rather similar to its original CF. The decrease of low cloud CF with the decrease of subsidence rate is better captured by the simulator CF than the original CF in the UKMO hadgem2-es.

For the spatial distributions of ω_{500}, the modeled fields tend to have a positive bias in ascending regimes but a negative bias in descending regimes, especially over southeast and equatorial Pacific. The negative biases near the inter-tropical convergence zone (ITCZ) are likely related to the common "double-ITCZ" problem manifested in precipitation. The pdfs of ω_{500} are broadly similar among the models and the ECMWF reanalysis.

Using the diagnostic framework discussed in Section 2.3, we decompose the differences between the model simulated and observed CWC and CF into the "large-scale (ω_{500}) error", "cloud parameterization error" and "co-variation error" (Figure 3b). A striking feature in Figure 3b is that the total errors (bottom panels) are predominantly contributed by the "cloud parameterization error", while the "large-scale error" and the "co-variation error" are relatively small and tend to cancel each other in most circulation regimes. This feature is even clearer in the tropical averaged errors (Figure 3c), with the contribution from the "cloud parameterization error" accounts for more than 95% of the tropical-averaged total errors. The relative small contribution by simulated large-scale vertical velocity to the total cloud error is not surprising, as ω_{500} is largely controlled by global energy balance, which is well constrained in the models. The dominance of the "cloud

parameterization errors" reflects our lack of understanding of the sub-grid scale cloud processes, which have been and will continue being the focus of ongoing model development.

Note that the actual values of model errors depend on which retrieval product is used and whether simulator CF is considered. For CWC, the sign of errors may change with response to CloudSat total or non-precipitating CWC. On the tropical average (Figure 3c), most model-simulated CWCs are close to the non-precipitating (lower bound) CWC in the UT above 400 hPa, but tend to be larger than the non-precipitating CWC and smaller than the total CWC in the boundary layer. In the middle troposphere between 800 and 500 hPa, most models overestimate CWC, even compared to CloudSat total CWC, except CSIRO, INM and UKMO. Considering the CALIPSO simulator CF for CNRM, MIROC, MRI and UKMO, the CF bias is mainly in the tropical tropopause layer (TTL) above 200 hPa and in the boundary layer, while the CF bias is relatively small for the rest of the troposphere. The CCCMA simulator CF bias is approximately opposite to that for the original CF. The IPSL simulator CF is closer to the observed in the TTL but deviates more from the observed at 200 hPa compared to the original CF, indicating potential problems with the simulator itself. The original model CF from CSIRO mk3.6 is very close to the observed at the pressure levels below 400 hPa. The similarity between CSIRO mk3.6 and UKMO hadgem2-es might because the convection schemes in both models are based on the mass flux convective parameterization by *Gregory and Rowntree* [1990], with varying degrees of adjustment.

Within the large-scale circulation (ω_{500}) regimes (Figure 3b), compared to CloudSat total CWC, the most common cloud parameterization errors in CWC are an underestimate of CWC in the UT and the boundary layer (note the sign change compared to non-precipitating CWC), but an overestimate of CWC in the middle troposphere and immediately near the surface. Two GISS models are different from the others in that their peak CWC is shifted too high in the UT, creating

a dipole of biases, with high (low) bias above (below) 350 hPa. Shown in Figure 3b are original CF biases, not the simulator CFs; thus, caution needs to be exercised for quantitative assessment. For the original CF, most models have high bias from 800 hPa up to 200 hPa, but low bias in the boundary layer. INM and MIROC have a high bias in CF near the surface, while MIROC has a low bias in CF in the UT. The opposite sign of high cloud CWC and CF biases for most models indicate the inconsistency in parameterized cloud mass and coverage, which is probably a result of model "tuning". Compared to other models, CSIRO mk3.6 and UKMO hadgem2-es have relatively small biases in the middle troposphere in terms of both CWC and CF.

3.1.3 Sorting clouds by LDV

Given the close correlation between ω_{500} and LDV, we expect clouds sorted by ω_{500} and LDV bear very similar features. A cursory look at Figure 4a suggests this is the case; however, a careful examination reveals interesting differences. First, the strongly lower-level convergent regimes (LDV < 0) are associated with stronger convective clouds in the lower to middle troposphere than the large-scale ascents (Figure 3a), both in the observation and models (except CSIRO mk3.6). This suggests that low and middle-level convection is strongly tired to the convergence between 1000 and 850 hPa, but less correlated with mid-level vertical velocity. Differences in the UT clouds are rather small when sorted by ω_{500} and LDV. Second, the low clouds are congregated within the divergent regimes of LDV between 0 and 5×10^{-6} s^{-1} for the observation and most models. CWC and CF are rather uniform within this range of LDV, unlike the obvious decreasing trends in them with decreasing subsidence rate (more negative ω_{500}). Hence, we think LDV is probably not the best quantity to characterize the variation of low cloud CWC and CF. The simulator CFs match better with the radar-lidar retrieved CF in the UT and in the boundary layer over the divergent regimes. In the two GISS models, some mid-to-high clouds emerge over highly

divergent regimes (the right most in the figure) which are relatively rare in occurrence frequency. These high clouds are not present in the observation or in the strongly descending regimes (Figure 3a). They indicate an inconsistency in the model simulations of LDV or ω_{500}.

The biases in simulated LDV across the models have similar patterns to the biases in ω_{500}. A common feature is that the models tend to be less polarized: there are positive (negative) biases in the regions of climatological convergence (divergence), resulting in less horizontal gradient in LDV.

We conduct the same decomposition for the three components of errors (Figures 4b and 4c). The gross features are very similar to those in Figures 3b and 3c. The cloud parameterization errors dominate the total errors, with the large-scale and co-variation errors approximately cancel each other. On the tropical average, the cloud parameterization error accounts for nearly all of the total errors.

3.2 Sorting clouds by thermodynamic variables SST, LTS and WVP

3.2.1 Correlation between SST, LTS and WVP

Figure 5 displays the scatter plots of LTS and WVP against SST. The linear correlation between SST and LTS is –0.73, with large scatter over cold SST less than 300 K. WVP bears an approximately exponential relationship with SST, with sharp increase of WVP when SST is greater than 300 K, related to the "super-greenhouse" effect of water vapor [*Raval and Ramanathan*, 1989]. The linear correlation between WVP and SST is 0.90. Such scatter plots are useful for us to understand the differences in cloud structures when sorted by SST, LTS and WVP separately.

3.2.2 Cloud profiles sorted by SST

In SST defined regimes, the general clustering of high (low) clouds in warm (cold) SST regimes is captured by all models (Figure 6a). Compared to the high clouds in the large-scale

ascending regimes, the magnitude of high clouds over SST warmer than 300 K appears to be weaker, which is because the cloud amounts in each bin of ω_{500} or SST is not weighted by the pdf of that bin. It is evident that all models, except GISS e2-h and e2-r, do not produce peak CWC as high in altitude as the CloudSat observation. Most models have maximum CWC around 500-600 hPa in the warm SST regimes, while IPSL cm5a and MRI cgcm3 peak even lower than 700 hPa. UKMO hadgem2-es exhibits one maximum in the UT resembling deep convective clouds, and one maximum in the middle troposphere reminiscent of cumulus congestus, but the CWC magnitude is very weak compared CloudSat total CWC (but comparable to CloudSat non-precipitating CWC), despite that the UKMO model includes snow, but no rain, in the CWC. In terms of the original CF, all models reproduce the large CF in the UT associated with convective detrainment, but many models have an overestimate of middle-level CF, which are obviously mitigated by the CALIPSO simulator CF, except in the MIROC miroc5. The MIROC high-resolution miroc4h (not shown) produces CWC very similar to the low-resolution miroc5 – the high bias of middle-level clouds in MIROC is not caused by model resolution.

In the cold SST (< 300 K) regimes, we observe an appreciable increase of cloud top height with increasing SST (not so much with the decrease of ω_{500} in Figure 3a), accompanied by the decrease of CWC and CF. In the models, the variations of cloud top height, CWC and CF with SST are not as linear as the observed. CCCMA, CSIRO, GFDL, and IPSL models reproduce the decrease of stratiform CF with increasing SST, but the increase of cloud top height is not as conspicuous as in the observed data. The UKMO simulator CF has a nice resemblance to the observed variation, although its original CF has little systematic change with SST. In other models, low cloud CF stays nearly constant or even increases with SST for SST lower than 295 K. As *Bony and Dufresne* [2005] pointed out, the variation of low cloud fraction or CWC with SST is very

important in determining the net cloud radiative forcing sensitivity to climate change. Hence, diligent efforts need to be expended to improve the model performance in this regard. Many models have clouds reaching above the freezing level in the very cold SST regimes around 290 K, likely caused by misplaced convective detrainments or intrusion of mid-latitude storms into the subtropics. Interestingly, such misplaced middle and high clouds in the models over the cold SST regimes do not occur in the large-scale subsidence regimes (Figure 3a), suggesting that these clouds are associated with wrongly simulated upward motions over climatologically cold waters. One region of problem may be the southeastern Pacific, where most models have negative ω_{500} bias, which may be responsible for the erroneous middle and high clouds there.

The differences of the coupled model simulated SST from the observed SST are generally small, except near the west coast of Peru and Chile, in the northwestern Pacific and north Atlantic. The biases are quite consistent across the models, indicating rather universal problems likely associated with biases in surface radiative and heat fluxes related to marine stratiform clouds.

Separating the modeled errors into three components in the regimes of SST (Figure 6b), we again find that the cloud parameterization errors dominate the total errors. However, the contributions of the errors due to inaccurate simulations of SST are non-negligible in each SST bin and on the tropical average (Figure 6c). For example, the departure of total errors from the parameterization errors for CWC are discernible at ~500 hPa for NCC, which are not present in the counterparts in the regimes of ω_{500}.

3.2.3 Cloud profiles sorted by LTS

Although LTS is strongly correlated with SST, the clouds sorted by LTS show some distinct characteristics (Figure 7a) that are not present in the SST sorted plots (Figure 6a), particularly for stratiform clouds. In CloudSat/CALIPSO observation, low cloud CWC and CF decrease linearly

with decreasing LTS until LTS < 12 K, along with an increase in cloud top height. This is similar to the changes of clouds when SST increases from less than 290 K to 300 K. However, although most models do not reproduce systematic changes of low clouds with SST, the majority of the models capture the increase of low clouds with increasing LTS, especially in the simulator CF. Only CNRM cm5 and two GISS models exhibit little or opposite changes of clouds with LTS in stable lower tropospheric regimes (LTS > 12 K). Such a difference in SST and LTS sorted cloud structure may tell us that the model simulated lower tropospheric temperature decouples from the underlying SST, while the observed lower tropospheric temperature is strongly tied to its lower boundary. The stratiform cloud parameterizations used in most of the CMIP5 models have ingredients, with varying extent, that mimic the observed relationship between LTS and low cloud fraction [*Klein and Hartmann*, 1993]. However, the decoupling of lower tropospheric temperature from SST is somewhat disturbing.

For LTS < 12 K, the sorted cloud structure is very similar to that in the regimes of SST > 300 K, confirming the strong correlation of LTS and SST over warm waters (Figure 5a). As shown in Figure 5a, more scatter is found between LTS and SST for SST < 300 K than for SST > 300 K. This is probably also related to why model simulated low clouds are well correlated with LTS when LTS > 12 K but not with SST < 300 K. The differences in the simulated LTS from the AIRS observation vary between models. The northeastern and southeastern Pacific near the coasts are two common regions of negative LTS biases. Positive biases are seen over the deep convective regions, including the western Pacific and equatorial ITCZ.

Figure 7b displays the three components of model errors in the functional space of LTS. Compared to the decomposed errors described before, we find that the relative contribution of "large-scale (LTS) error" is significantly larger than the large-scale errors associated with ω_{500},

LDV, or SST, for both high and low clouds. For example, CCCMA canesm2 simulated LTS has an approximately uniform high bias over the tropical ocean, causing the pdf of LTS shifted towards higher stability values. This is associated with a negative CWC bias in the relatively unstable regimes (LTS < 15 K) and a positive CWC bias in the stable regimes (LTS > 15 K), compared to CloudSat total CWC. For LTS between 12 and 15 K, the negative CWC bias due to the imperfect LTS is superimposed on the dipole pattern (negative above, positive below) of cloud parameterization errors, resulting in an amplification of the negative UT CWC bias, approximate cancellation of errors between 600 and 700 hPa, and a reversal of positive to negative cloud bias below 700 hPa. Over this regime (12 < LTS < 15 K), the large-scale error is not negligible and the co-variation error amplifies the large-scale error below 600 hPa, instead of compensating it. Similarly, the error associated with the LTS simulation in the GFDL cm3 also makes an appreciable departure of the total error from the cloud parameterization error in each LTS bin and on the tropical average (Figure 7c). Even comparing to the non-precipitating CWC, the "large-scale (LTS) error" is more pronounced than those associated with ω_{500}, LDV and SST.

3.2.4 Cloud profiles sorted by WVP

The clouds sorted by WVP (Figure 8a) exhibit very similar structure to those sorted by SST. It is related to the close correlation between SST and WVP shown in Figure 5b. The only difference may be in the erroneous middle and high clouds over the coldest SST regimes, which do not show up in the lowest values of WVP. This might be related to the fact that many models overestimate WVP over climatologically cold waters (see corresponding maps of ΔWVP, especially for GISS and MRI). The inconsistent patterns of sorted clouds in different large-scale regimes reveal certain biases in the model simulated large-scale parameters. Most models underestimate WVP over

climatologically warm waters and large-scale ascending regions. This may cause the underestimate of CWC over these regions.

For the relative contributions of three error components, we find that the errors associated with the simulated WVP can be of comparable magnitude to that of the cloud parameterization error in individual WVP bins (Figure 8b). On the tropical average (Figure 8c), the deviations of the total errors from the parameterization errors are evident in a number of models, although the dominance of the cloud parameterization errors still holds universally. Using GISS e2-h as an example, the dry bias over the very moist regions (WVP > 50 mm) creates a negative CWC bias (Figure 8b), outweighing the positive CWC bias caused by the parameterization error in the UT and in the lower troposphere. The total CWC error in this regime resembles that from the large-scale error for WVP > 50 mm. The UKMO hadgem2-es has a widespread dry bias over the tropical ocean, which is partially responsible for the negative CWC bias over the moist areas with WVP > 50 mm.

3.3 Sorting clouds by relative humidity

RH has been used commonly to parameterize CF [e.g., *Gettelman et al*., 2010]. Thus, examining the cloud profiles as a function of RH is directly relevant to validate the cloud parameterizations. Figure 9a shows the CWC and CF profiles sorted by RH at corresponding levels (MIROC miroc5 did not provide RH output on the ESG). For the observed data, we use ECMWF RH, as AIRS RH data are missing over thick cloudy regions. Note that the RH data from the two GISS models are with respect to (w.r.t.) water throughout the troposphere, unlike other models and the ECMWF reanalysis, in which the saturation vapor pressure is with respect to ice (water) above (below) the freezing level. For the ECMWF reanalysis, RH is with respect to water above 273.16 K, to ice below 250.16 K, and with respect to a mixture of both in between (http://www.ecmwf.int/publications/manuals/metview/manual/Relative_Humidity.html). The exact

temperature threshold for the switch from w.r.t. water to ice in the CMIP5 models may be quite different. Hence, there is an ambiguity in comparing RH from different models and the reanalysis. A rescaling of RH with a unified temperature threshold across the models may be needed for direct comparison of modeled RH profiles. Nevertheless, the standard model outputs of RH on the ESG are used in Figure 9a.

In general, there is a dominant peak of CWC in the UT associated with RH greater than 60%, which is true for several models such as BCC csm1, GFDL cm3, IPSL cm5a, NCC noresm, and to a lesser extent, CSIRO mk3.6 and UKMO hadgem2-es. CCCMA canesm2 seems to have two CWC maxima, a weaker one in the UT and a stronger one in the middle troposphere associated with high RH. CNRM cm5, MIROC miroc4h, MRI cgcm3 have rather uniform distribution of CWC in the vertical over the regimes of RH > 60%. GISS e2-h and e2-r capture the correspondence of high RH with high clouds, despite that their RH values are lower than other models in the UT because they are w.r.t. to water.

The variations of low clouds with RH differ substantially in the models from the observations. The most notable discrepancy is that the models do not produce sufficient CWC in the areas of relative humidity less than 40%, while the observed CloudSat CWC spans the full range of RH, even when RH is less than 20%. Despite that the potential high bias of CloudSat CWC for thin liquid clouds, the missing clouds in the dry regions in the models may still be an important issue that requires further study. Applying the simulator CFs yields better agreement with the observation than the original CFs for most of the cloud features.

For the three error components, besides the dominance of the cloud parameterization errors, we observe significant contribution of the large-scale error in each RH bin (Figure 9b). The underestimate of high RH by the two GISS models apparently contributes to the negative (positive)

CWC bias in the UT over relatively moist (dry) regions with RH greater than (less than) 60%. On the tropical average (Figure 9c), the departures of total errors from the cloud parameterization errors are evident for a number of models.

3.4 Sorting clouds by precipitation

Since precipitation is not a large-scale environmental parameter that is used to parameterize clouds, the purpose of sorting cloud profiles by precipitation is to visualize their co-variations as the two are closely coupled. Figure 10a shows that the difference between the models and the observation is quite large, while the vertical structures of precipitating clouds are rather similar among the models. Many models exhibit a tri-modal structure of clouds [*Johnson et al.*, 1999], with increasing cloud top heights with stronger precipitation. Most models do not produce clouds in the very weak precipitation regimes (rain rate < 0.05 mm/day), except CSIRO mk3.6, MRI cgcm3, GISS e2-h and e2-r (which produce middle-level clouds over the light rain regimes), while the CloudSat/CALIPSO observation shows significant CWC and CF in the lightly precipitating regimes. *Haynes and Stephens* [2007] reported that the occurrence frequency of light rain observed by CloudSat is higher than previously thought and that simulated by climate models. The simulator CFs exhibit a two-modal structure of clouds, closer to the observation than the original CFs, although differences from the observed CF are still quite evident.

We conduct a similar decomposition of three error components for clouds in the precipitation regimes. Although we do not regard them as the sources of cloud errors, they do tell us how much precipitation errors are associated with cloud errors (Figure 10b). Again, the errors of parameterized clouds in a given precipitation regime are predominant in the total errors, while the common overestimate of precipitation rates in the models translate into positive CWC biases, partially compensating the negative parameterization biases in most models. For the two GISS

models, the positive CWC biases associated with the cloud parameterization errors are exacerbated by the positive biases associated with the overestimate of high precipitation rates (see Figures 10b and 10c).

4. Conclusions

This paper examines the cloud distributions in large-scale parameter regimes, focusing on the cloud vertical structures and their variations with large-scale variables. The regime-dependent model biases from the observations are decomposed into three components, the large-scale errors, cloud parameterization errors and the co-variation errors. Despite of a variety of discrepancies in the simulated cloud structures, a universal feature is that in all models, the cloud parameterization errors dominate, while the large-scale and the co-variation errors are secondary. This finding confirms the deficiency in the current state of knowledge about the governing mechanisms for sub-grid cloud processes, as pointed out in the IPCC AR4. Efforts should be devoted to ongoing improvements of cloud parameterization schemes. On the other hand, the relatively small contribution of large-scale errors to the total cloud errors suggests that all CMIP5 models produce reasonable large-scale "mean" states, which are arguably the first-order depiction of current climate. This is comforting as these large-scale state variables are constrained by global energy balance, such that modelers usually have less freedom in "tuning" them than modifying cloud parameterizations. With the new satellite observations as observational metrics, we envision that significant progress will be made in cloud parameterizations in the coming years.

This study focuses on comparing model simulated cloud water content (CWC) and cloud fraction (CF) vertical structures with 13 atmosphere-ocean coupled models. We recognize that large uncertainties exist in the quantitative assessment of model simulated CWC and CF. For CWC, a major uncertainty is associated with the inclusion of precipitating condensates in CloudSat CWC

retrieval, while CMIP5 models have non-uniform treatments for snow and rain. This makes a fair comparison extremely difficult. Using the total CWC and non-precipitating CWC as the upper and lower bounds of observed CWC, we find that the signs of model biases in CWC, relative to the two bounds, reverse for many models, suggesting that the modeled values are actually within the observed uncertainties. Some models exhibit persistent sign of errors with respect to both CWC quantities, indicating a true signal of biases. For example, the underestimate of CWC by INM cm4 and the overestimate of CWC by the two GISS models are robust at most of the vertical levels. Many models do not include rain water in the liquid water content output. Therefore, compared to the non-precipitating CWC, an overestimate of LWC is prevailing. However, compared to the total CWC, most models have an underestimate of LWC. In the UT, many models produce CWC values close to CloudSat non-precipitating CWC, but much less than the total CWC.

For cloud fraction, we find that using CALIPSO simulator CF generally brings the model results closer to CloudSat/CALIPSO combined CF. One of the best examples is UKMO hadgem2-es: the simulator CF shows a much better agreement with the observed CF for both high and low clouds. Hence, we think it is beneficial to apply the simulator approach in model evaluations, especially in terms of CF, for which instrument sensitivity has a large impact on the observed values. Besides CALIPSO simulator, CloudSat, ISCCP and other simulators will also be useful. Using CALIPSO simulator CF clearly mitigates the seemingly high bias of middle-to-high cloud CF shown in the original CF by many models; however, the biases of CF in the UT above 200 hPa and in the boundary layer are still present, for either the simulator CF or the original CF, suggesting that improvements in the model simulated CF are still needed.

For the large-scale errors, relatively larger errors are associated with thermodynamic parameters such as LTS, WVP and RH, than dynamic parameters such as ω_{500} and LDV. The

model simulated temperature and moisture structures are influenced by diabatic heating resulting from latent heat release and radiation. The latter are coupled with convection and clouds. Hence, the relatively large errors associated with LTS, WVP and RH than with ω_{500} and LDV may be a manifestation of the feedbacks of cloud errors onto the large-scale thermodynamic parameters. The model improvements for clouds are certainly a "two-way" game – the coupling between clouds and large-scale fields requires diligent validations of both fields.

In this study, we show that the model simulated variations of clouds with height bear large discrepancies from the observations. The model simulated deep convective clouds (in terms of CWC) usually do not penetrate as high into the UT as the observed, although the maximum detrainment height shown by CF is correctly placed around 200 hPa. The inconsistency between CWC and CF vertical structures indicate the mismatch in cloud parameterizations. For low clouds, we find that the change of cloud top height with large-scale parameters is usually missing in the models. The best large-scale parameter in the models that characterizes the changes of low cloud properties (CWC, CF and cloud top height) is LTS. Other large-scale variables have little bearing on the low cloud amounts, while their counterparts from the observations show clear correlation with the systematic low cloud amount changes. For climate sensitivity, such variations of cloud amount with large-scale parameter are probably more relevant than the magnitudes of mean cloud amounts. This is an area that significant improvements are needed for next generation of cloud parameterizations.

The large-scale parameters used for this study are not a complete list of variables that may affect cloud simulations in climate models. Additional parameters, such as wind profiles (including wind shear) and aerosols may be also important. Furthermore, the combinations of large-scale parameters and their joint distributions may reveal additional useful information for physical

mechanisms that control cloud variabilities. Continued investigations using conditional sampling approach with innovative combination of various large-scale quantities may lead to potential breakthrough in cloud parameterizations.

Acknowledgments. We thank the funding support from NASA COUND and AST, the CloudSat/CALIPSO mission teams and relevant satellite mission projects. We acknowledge PCMDI for archiving CMIP5 model simulations and the ECMWF Data Server for ECMWF Interim Reanalysis data. This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

References

Austin, R. T., A. J. Heymsfield, and G. L. Stephens (2009), Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature, *J. Geophys. Research,* 114, D00A23, doi:10.1029/2008JD010049, 2009.

Bodas-Salcedo, A., M. J. Webb, M. E. Brooks, M. A. Ringer, K. D. Williams, S. F. Milton, and D. R. Wilson (2008), Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities, *J. Geophys. Res*., 113, D00A13, doi:10.1029/2007JD009620.

Bodas-Salcedo, A. et al, (2011): COSP: Satellite simulation software for model assessment. *Bull. Amer. Met. Soc*., 92, 1023-1043. doi:10.1175/2011BAMS2856.1.

Bennhold, F., and S. Sherwood (2008), Erroneous Relationships among Humidity and Cloud Forcing Variables in Three Global Climate Models. *J. Climate*, 21, 4190-4206.

Bodas－Salcedo A., M. J. Webb, M. E. Brooks, M. A. Ringer, K. D. Williams, S. F. Milton, and D. R. Wilson (2008), Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities, *J. Geophys. Res*., 113, D00A13, doi:10.1029/2007JD009620, 2008.

Bony, S., J.-L. Dufresne, H. LeTreut, J.-J. Morcrette, and C. Senior, 2004: On dynamic and thermodynamic components of cloud changes. *Climate Dyn*., **22**, 71-86.

Bony S., J.-L. Dufresne, 2005, Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, *Geophys. Res. Lett*., 32, L20806, doi:10.1029/2005GL023851.

Gettelman, A., X. Liu, S. J. Ghan, H. Morrison, S. Park, and A. J. Conley (2010) Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the

Community Atmospheric Model. *Journal of Geophysical Research*, **115**, D18216, doi:10.1029/2009JD013797.

Gregory, D. and P. R. Rowntree, (1990): A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure, *Mon. Weather Rev.*, 118, 1483-1506.

Haynes, J. M., and G. L. Stephens (2007), Tropical oceanic cloudiness and the incidence of precipitation: Early results from CloudSat, *Geophys. Res. Lett.*, 34, L09811, doi:10.1029/2007GL029335.

Huang, X.L., V. Ramaswamy, and M.Daniel Schwarzkopf (2006), Quantification of the source of errors in AM2 simulated tropical clear-sky outgoing longwave radiation, *J. Geophys. Res*., 111, D14107, doi:10.1029/2005JD006576.

Jiang, J. H., H. Su, C. Zhai, and V. Perun, et al., Evaluation of Cloud and Water Vapor Simulations in CMIP5 Climate Models Using NASA "A-Train" Satellite Observations, *J. Geophys. Res*, 117, D14105, doi:2011JD017237, 2012.

Johnson, R. H., T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert (1999), Trimodal characteristics of tropical convection, *J. Clim*., 12, 2397–2418.

Klein, S. A., and D. L. Hartmann (1993), The seasonal cycle of low stratiform clouds, *J. Clim*., 6, 1587– 1606.

Randall, D. A., et al. (2007), Climate models and their evaluations, in *Climate Change 2007: The Physical Sciences Basis, Contribution of Working Group I to the Fouth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., chapter 8, 589-662, Cambridge Univ. Press, U.K.

Raval, A., and V. Ramanathan (1989), Observational determination of the greenhouse effect, *Nature*, 342, 758– 762.

Su, H., J. H. Jiang, D.G. Vane, and G.L. Stephens, Observed Vertical Structure of Tropical Oceanic Clouds Sorted in Large-scale Regimes, *Geophys. Res. Lett.*, *35*, doi:10.1029/2008GL035888, 2008.

Su, H., J.H. Jiang, J. Teixeira, A. Gettelman, X. Huang, G. Stephens, D. Vane, and V.S. Perun, Comparison of Regime-Sorted Tropical Cloud Profiles Observed by CloudSat with GEOS5 Analyses and Two General Circulation Model Simulations, *J. Geophys. Res. 116*, D0910, doi:10.1029/2010JD014971, 2011.

Taylor, K.E., R.J. Stouffer, G.A. Meehl: An Overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc*., 93, doi:10.1175/BAMS-D-11-00094.1., 2012.

Waliser, D.E., et al. (2009), Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.* 114, D00A21, doi:10.1029/2008JD010015.

Williams K. D., M. A. Ringer, C. A. Senior, 2003: Evaluating the cloud response to climate change and current climate variability. *Clim Dyn*, 20, 705–721.

Wyant, M. C., C. S. Bretherton, J. T. Bacmeister, J. T. Kiehl, I. M. Held, M. Zhao, S. A. Klein, and B. A. Soden, 2006: A comparison of low-latitude cloud properties and their response to climate change in three AGCMs sorted into regimes using mid-tropospheric vertical velocity. *Climate Dyn*., 27, 261-279.

Zhang, Y., et al. (2007), Cluster analysis of tropical clouds using CloudSat data, *Geophys. Res. Lett*., **34**, L12813, doi:10.1029/2007GL029336.

Figure Captions

Figure 1. Annual zonal-mean (top) original and (middle) CALIPSO simulator cloud fraction, and (bottom) tropical mean (30ºS-30ºN) original and CALIPSO simulator CFs from six CMIP5 models. Model monthly outputs from 1980-2004 are used.

Figure 2. Scatter plot of lower-level divergence averaged between 1000 and 850 hPa against mid-tropospheric vertical pressure velocity (ω_{500}) for all tropical oceanic grid boxes. Both quantities are from ECMWF interim reanalysis averaged from January 1980 to December 2004.

Figure 3. (a) Vertical profiles of cloud water content (CWC, in color shadings) and cloud fraction (CF, original CF in black contours, CALIPSO simulator CF in white contours) sorted by ω_{500} for the CloudSat/CALIPSO retrievals and the CMIP5 atmosphere-ocean coupled model simulations. The PDFs of ω_{500} are shown in gray dashed lines. For the observation, the ECMWF reanalysis ω_{500} is used to sort CloudSat total CWC and CloudSat/CALIPSO combined CF. The difference of model simulated ω_{500} from the ECMWF reanalysis ω_{500} is shown for each model.

Figure 3. (**b)** Decomposition of model simulated CWC (in color shadings) and CF (in contours) errors in the regimes of ω_{500}: (top row) the large-scale error, (second row) the cloud parameterization error, (third) the co-variation error, and (fourth) the total cloud errors. The CWC errors are relative to the CloudSat total CWC. The model original CFs are used. Only 12 models are shown due to space constraint. GISS e2-r is not shown and similar to e2-h.

Figure 3. (**c)** Tropical averages of model simulated CWC and CF errors and the contributions of three error components. The blue (green) curves are the CWC errors relative to the CloudSat total (non-precipitating) CWC. The red (black) curves are the original (CALIPSO simulator) CF errors relative to the CloudSat/CALIPSO CF retrieval. Only 12 models are shown due to space constraint. GISS e2-r is not shown and is similar to e2-h.

Figure 4. Same as Figure 3, except lower-level divergence (LDV) is used to sort cloud profiles. The errors relative to the CloudSat non-precipitating CWC are not shown, and the CALIPSO simulator CF errors are not shown.

Figure 5. Annual mean lower-tropospheric stability (LTS) calculated from AIRS temperature data and AMSR-E water vapor path (WVP) scattered against ECMWF sea surface temperature (SST) for all tropical oceanic grid boxes. The monthly AIRS data from September 2002 to May 2011 and the AMSR-E data from June 2002 to December 2010 are used.

Figure 6. Same as Figure 3, except SST is used to sort cloud profiles.

Figure 7. Same as Figure 4, except LTS is used to sort cloud profiles.

Figure 8. Same as Figure 4, except WVP is used to sort cloud profiles.

Figure 9. Same as Figure 4, except relative humidity (RH) is used to sort cloud profiles. The maps of 3-dimensional RH are not shown due to space constraint. MIROC miroc4h (with only CWC, not CF) is shown instead of MIROC miroc5 because RH data are not available for miroc5 and CF is not available for miroc4h.

Figure 10. Same as Figure 4, except surface precipitation is used to sort cloud profiles.

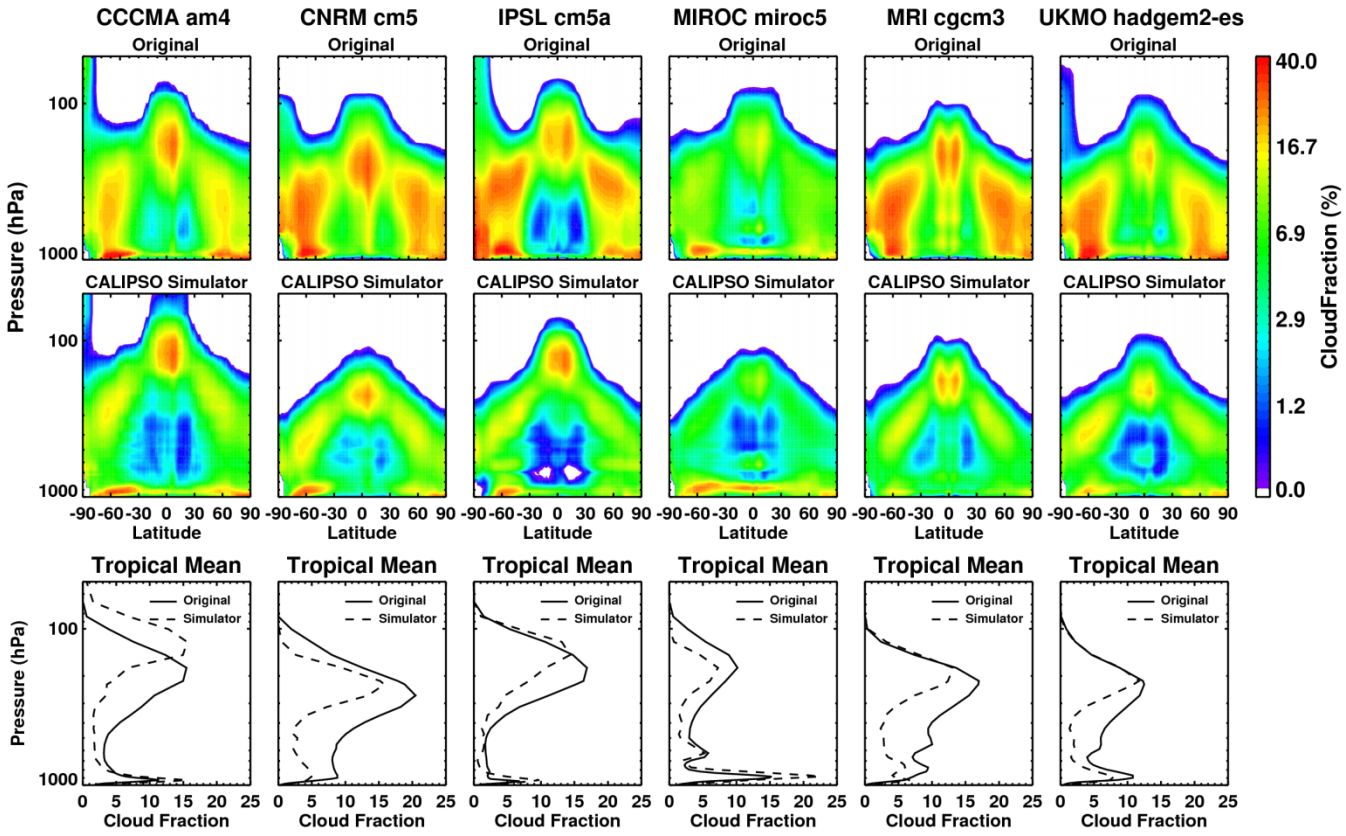
CCCMA am4
CNRM cm5
IPSL cm5a
MIROC miroc5
MRI cgcm3
UKMO hadgem2-es
Original
CALIPSO Simulator
Pressure (hPa)
100
1000
-90 -60 -30 0 30 60 90
Latitude
CloudFraction (%)
40.0
16.7
6.9
2.9
1.2
0.0
Tropical Mean
Original
Simulator
0 5 10 15 20 25
Cloud Fraction

Figure 1

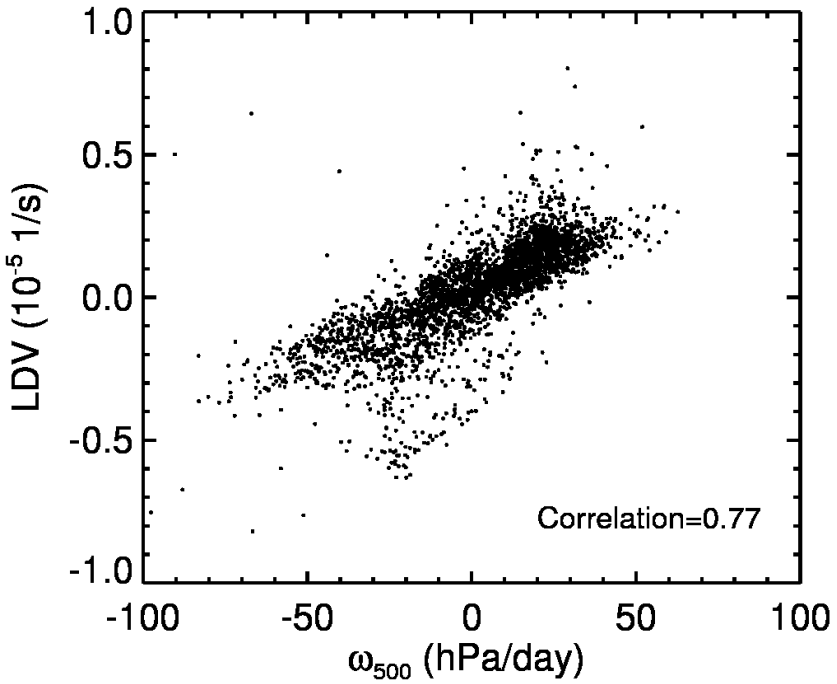
1.0
0.5
0.0
-0.5
-1.0
LDV (10^{-5} 1/s)
-100
-50
0
50
100
ω_{500} (hPa/day)
Correlation=0.77

Figure 2

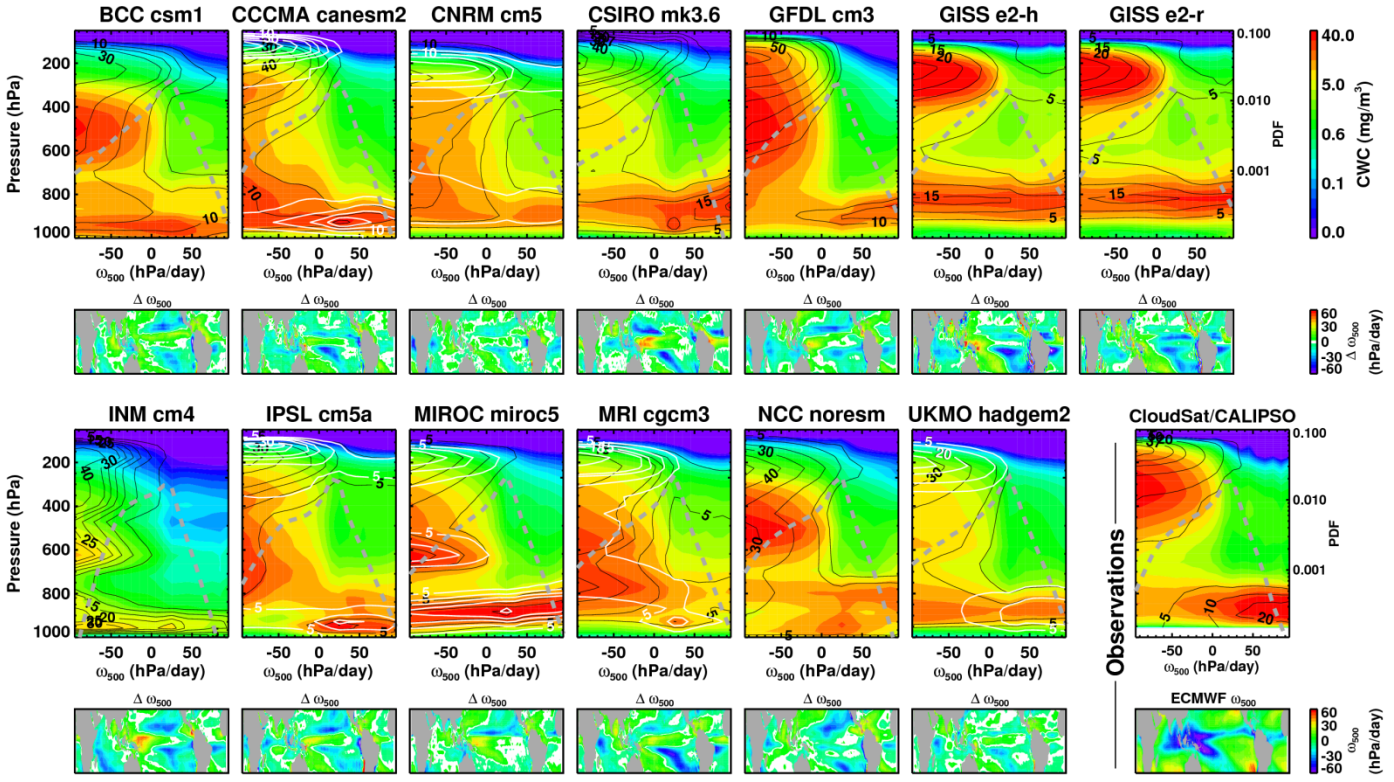
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
ECMWF ω500
Observations
Pressure (hPa)
ω500 (hPa/day)
Δ ω500
PDF
CWC (mg/m3)

Figure 3a

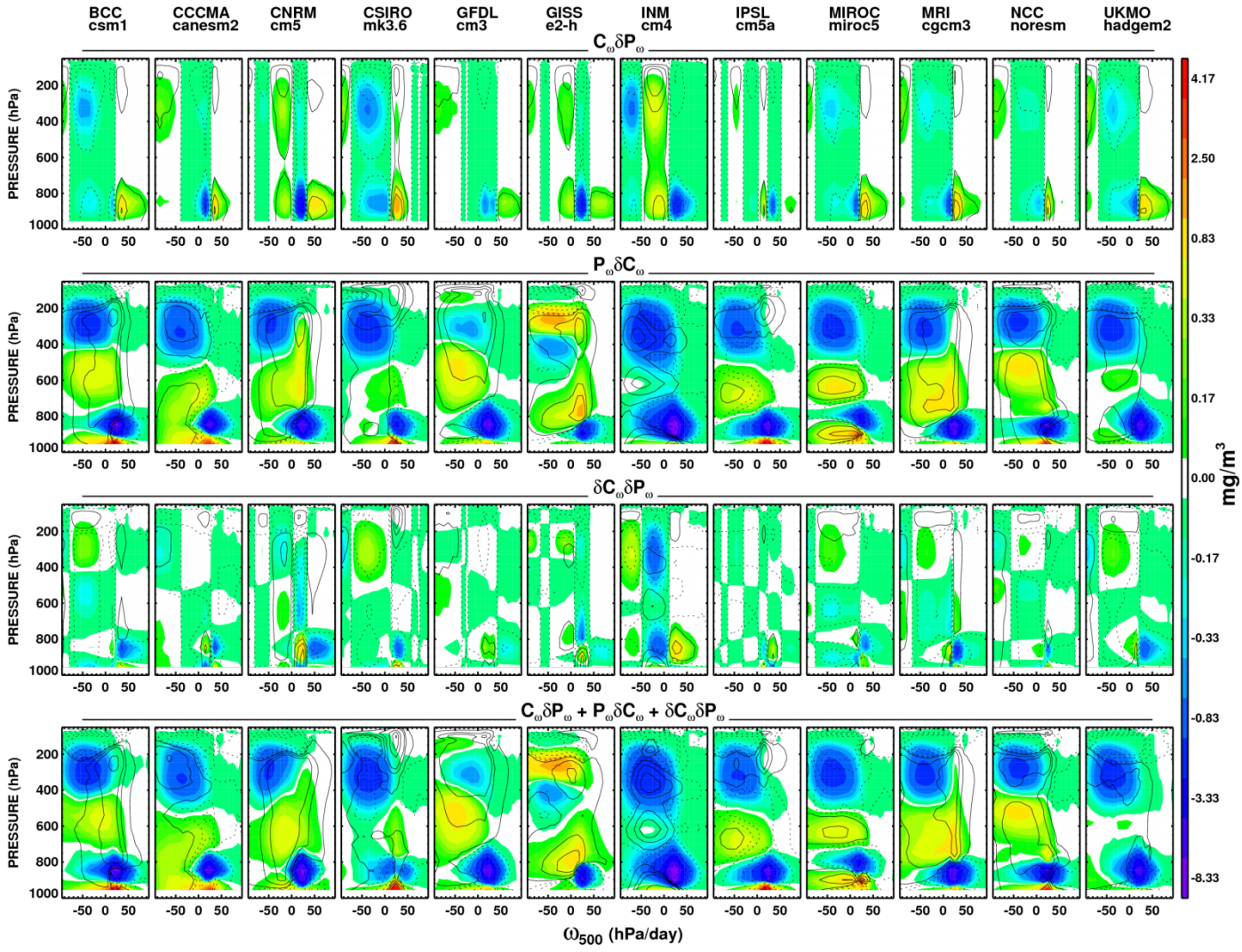
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_\omega\delta P_\omega$
$P_\omega\delta C_\omega$
$\delta C_\omega\delta P_\omega$
$C_\omega\delta P_\omega + P_\omega\delta C_\omega + \delta C_\omega\delta P_\omega$
PRESSURE (hPa)
200
400
600
800
1000
-50 0 50
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3
ω_{500} (hPa/day)

Figure 3b

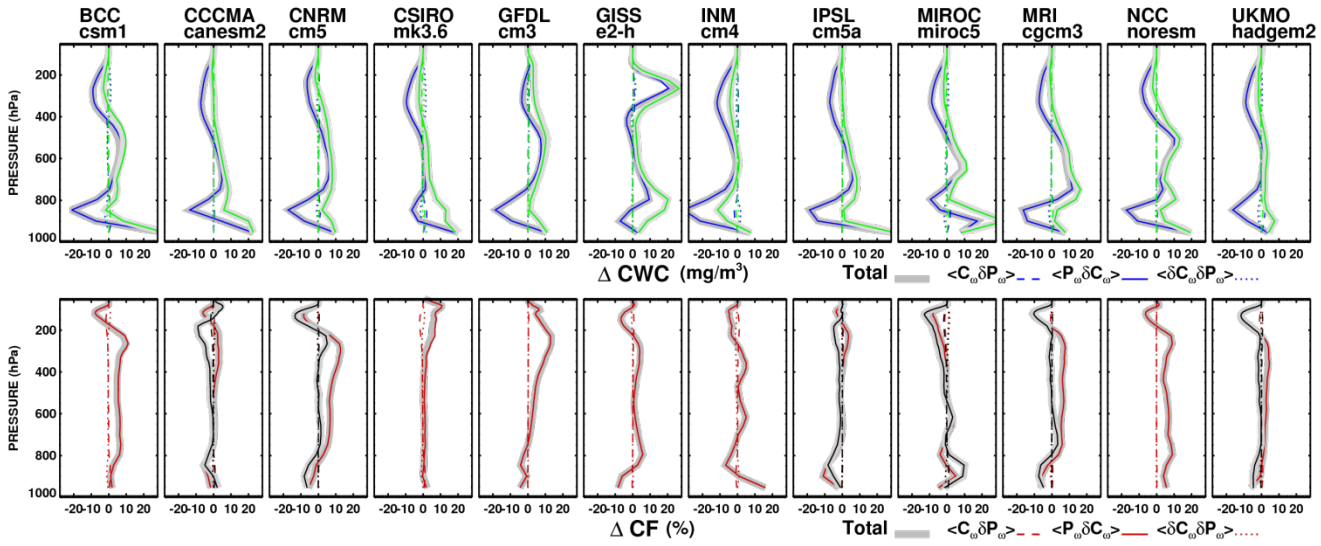
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<Cω δPω>
<Pω δCω>
<δCω δPω>
Δ CF (%)

Figure 3c

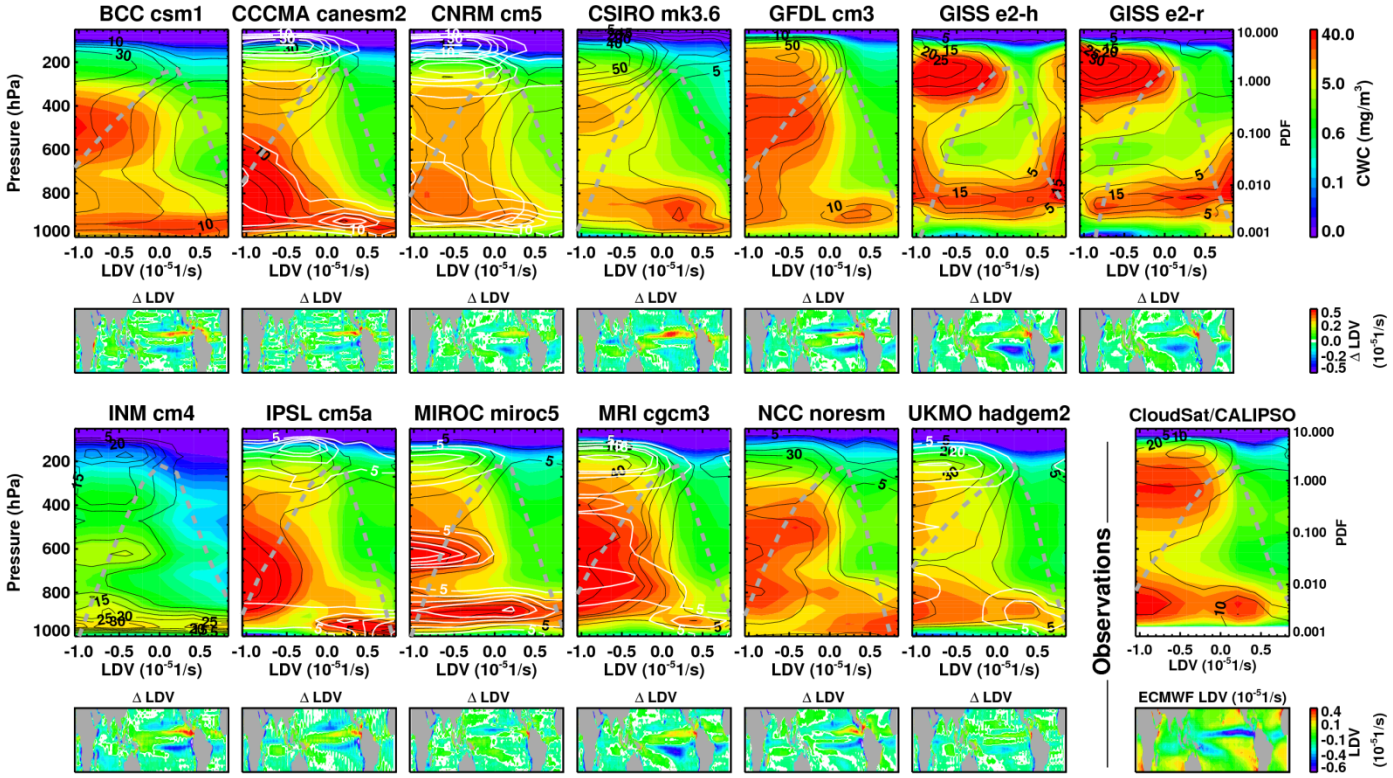
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
Observations
Pressure (hPa)
LDV (10^-5 1/s)
Δ LDV
PDF
CWC (mg/m^3)
ECMWF LDV (10^-5 1/s)

Figure 4a

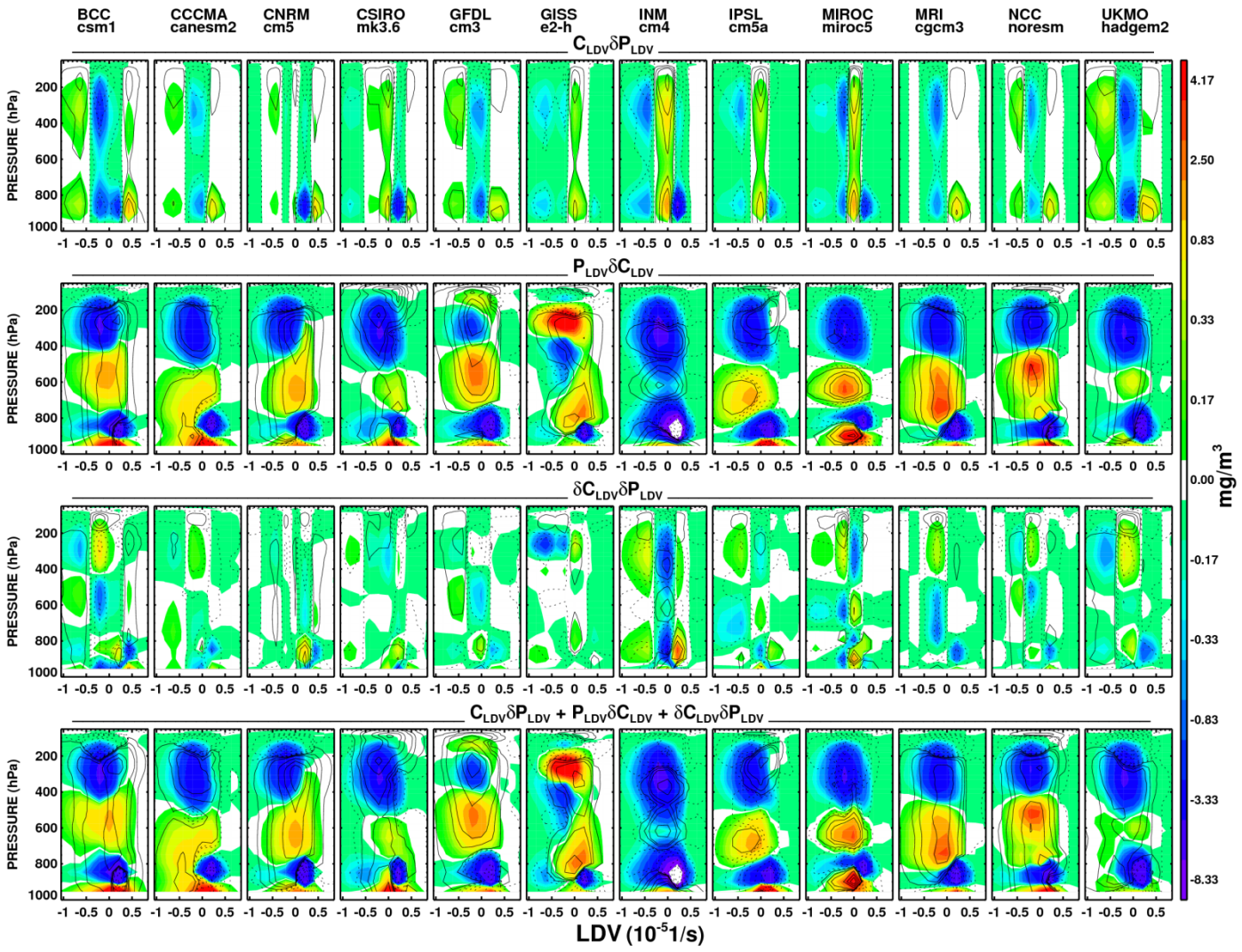
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_{LDV}\delta P_{LDV}$
$P_{LDV}\delta C_{LDV}$
$\delta C_{LDV}\delta P_{LDV}$
$C_{LDV}\delta P_{LDV} + P_{LDV}\delta C_{LDV} + \delta C_{LDV}\delta P_{LDV}$
PRESSURE (hPa)
200
400
600
800
1000
-1 -0.5 0 0.5
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3
LDV (10^{-5}1/s)

Figure 4b

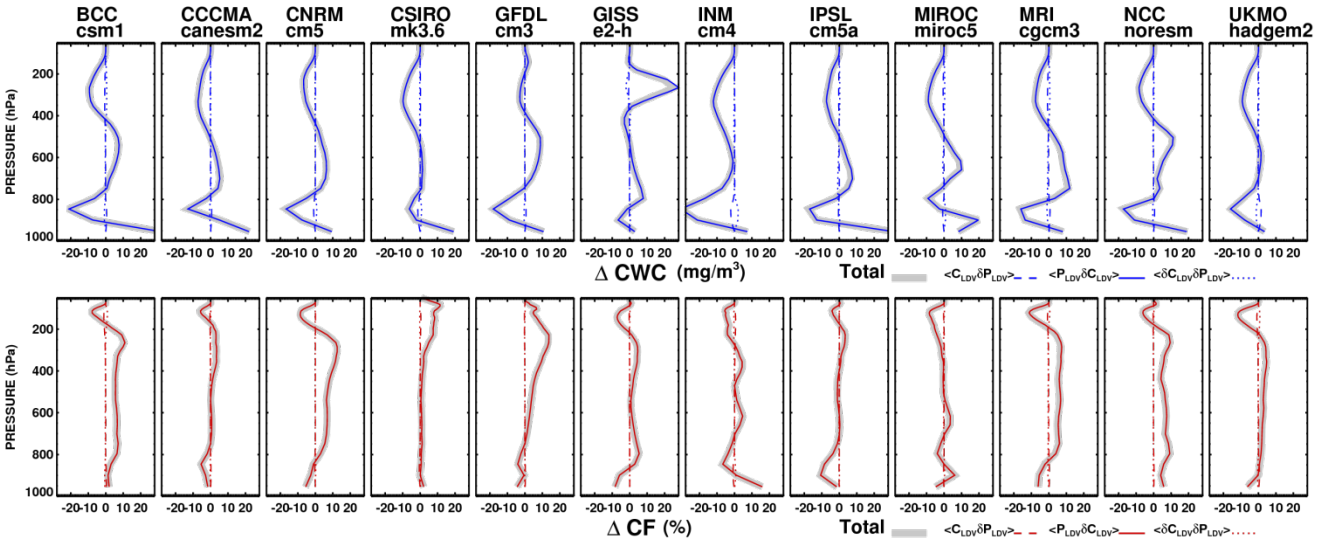
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<CLDVδPLDV>
<PLDVδCLDV>
<δCLDVδPLDV>
Δ CF (%)

Figure 4c

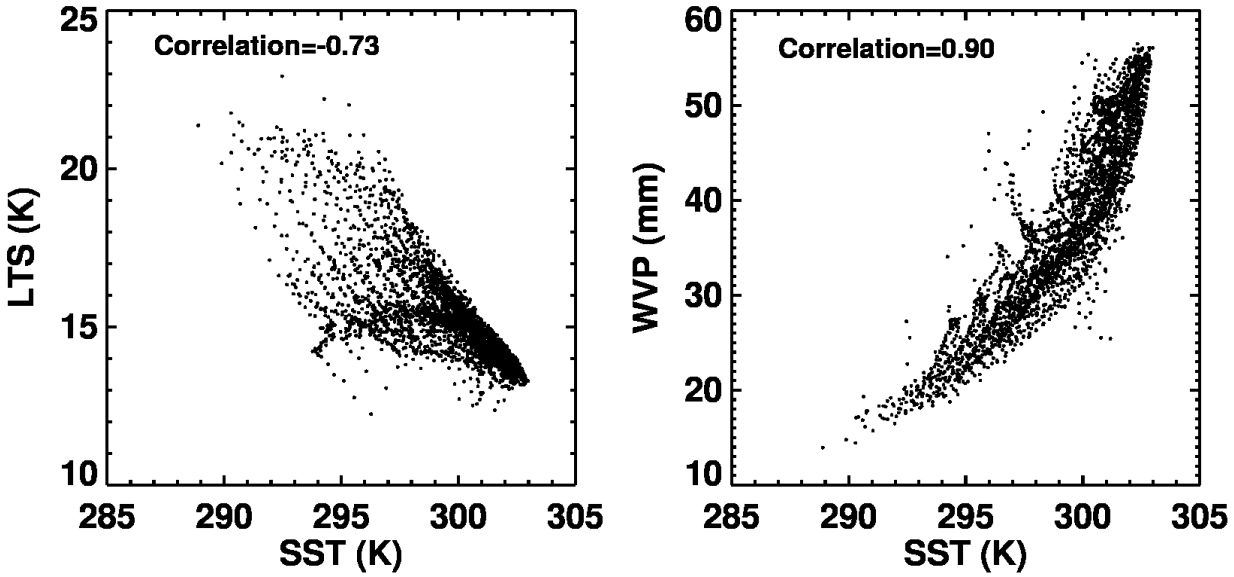
Correlation=-0.73
LTS (K)
SST (K)
25
20
15
10
285
290
295
300
305
Correlation=0.90
WVP (mm)
SST (K)
60
50
40
30
20
10
285
290
295
300
305

Figure 5

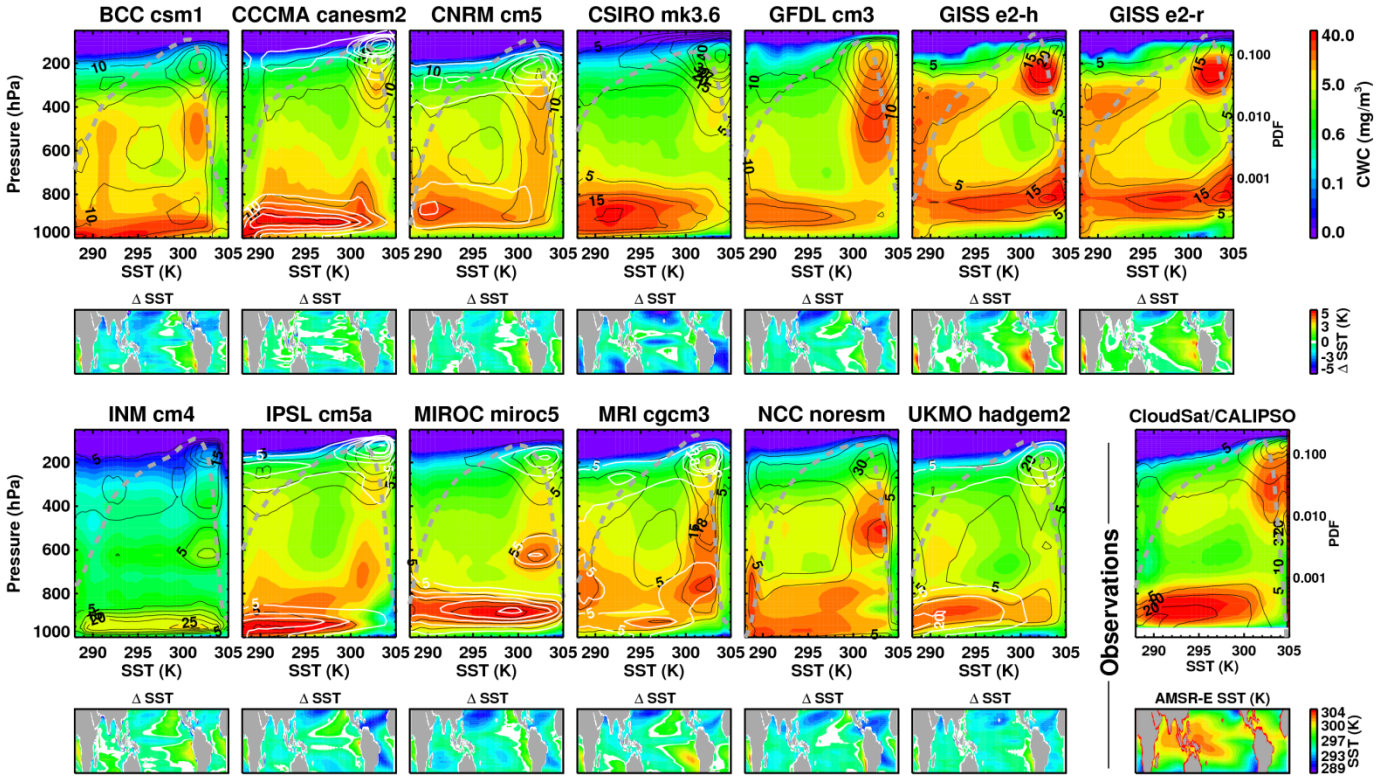
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
Observations
Pressure (hPa)
SST (K)
Δ SST
PDF
CWC (mg/m3)
AMSR-E SST (K)

Figure 6a

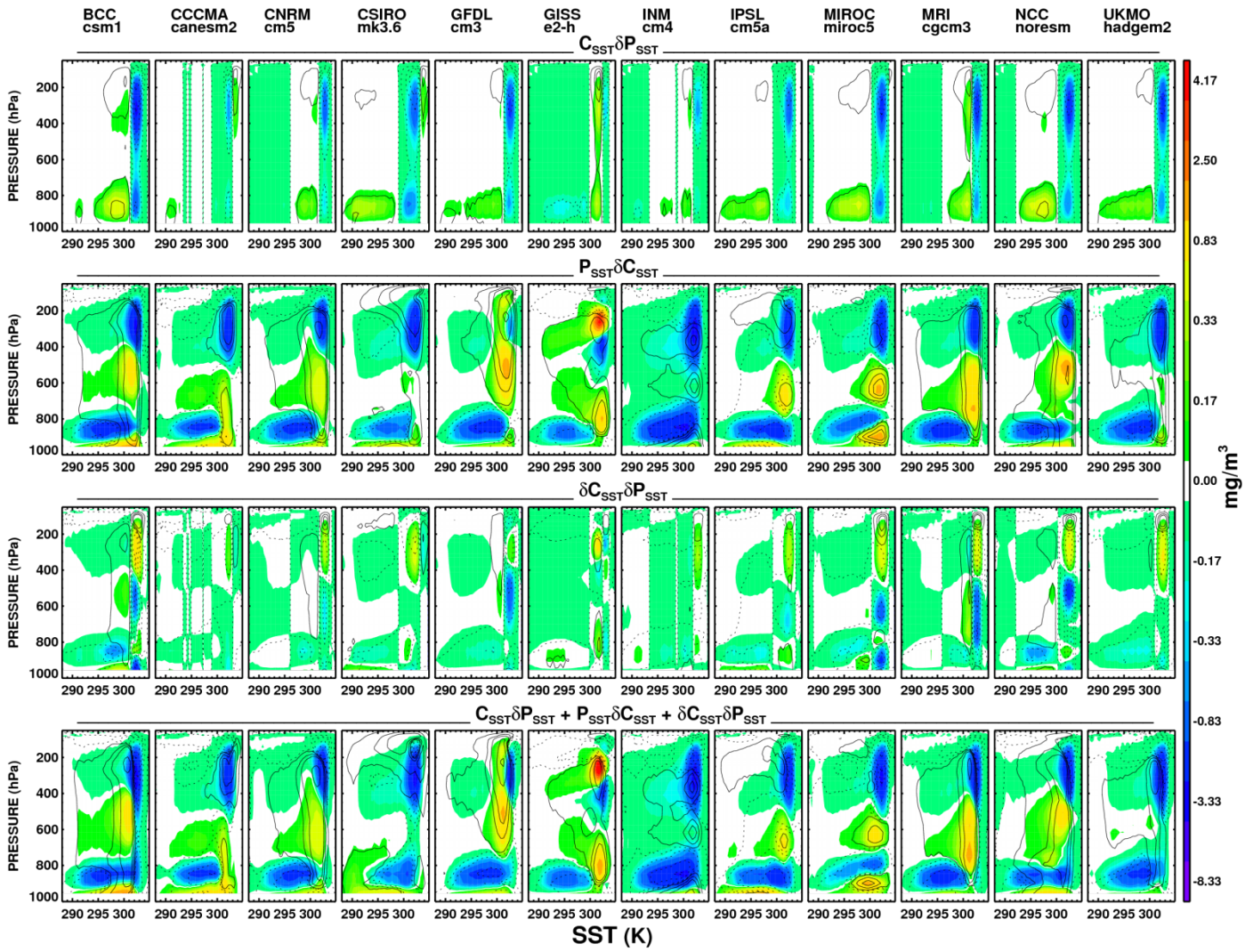
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_{SST}\delta P_{SST}$
$P_{SST}\delta C_{SST}$
$\delta C_{SST}\delta P_{SST}$
$C_{SST}\delta P_{SST} + P_{SST}\delta C_{SST} + \delta C_{SST}\delta P_{SST}$
PRESSURE (hPa)
200
400
600
800
1000
290 295 300
SST (K)
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3

Figure 6b

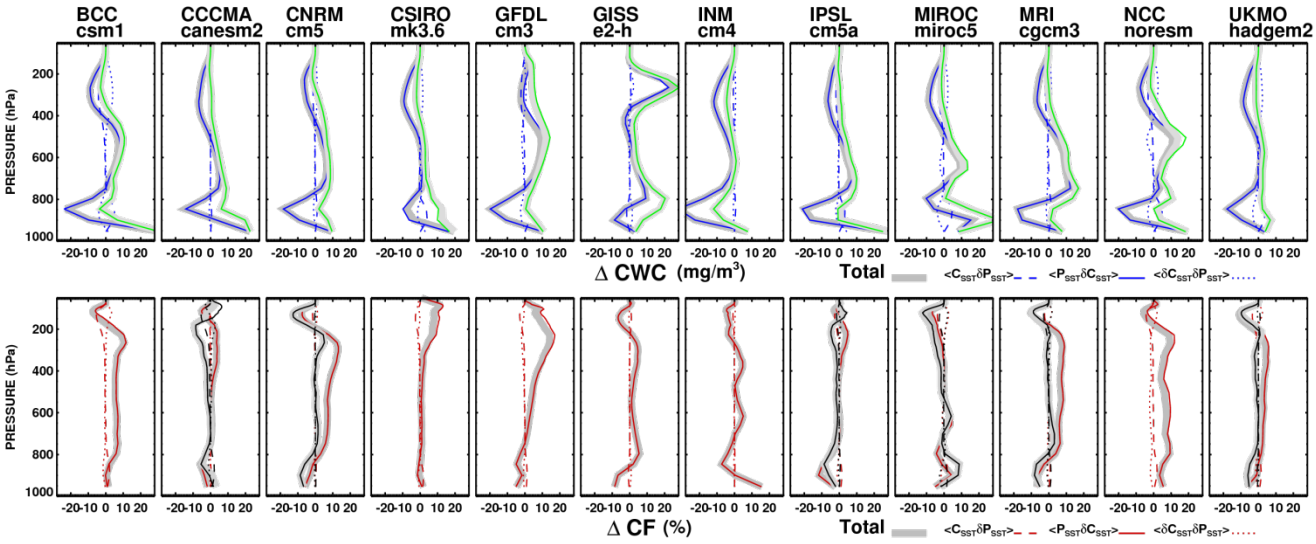
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<CSSTδPSST>
<PSSTδCSST>
<δCSSTδPSST>
Δ CF (%)

Figure 6c

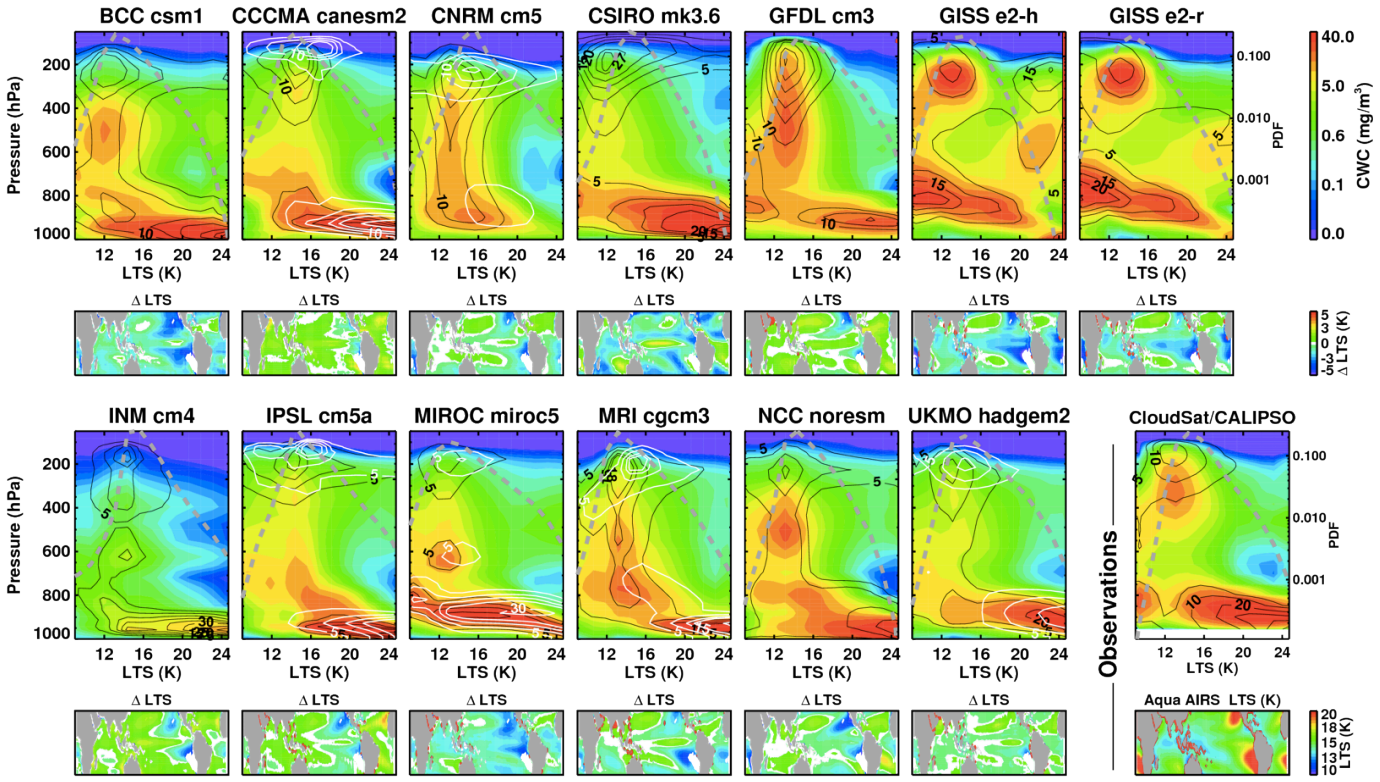
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
Aqua AIRS LTS (K)
Observations
Pressure (hPa)
LTS (K)
Δ LTS
PDF
CWC (mg/m3)
Δ LTS (K)

Figure 7a

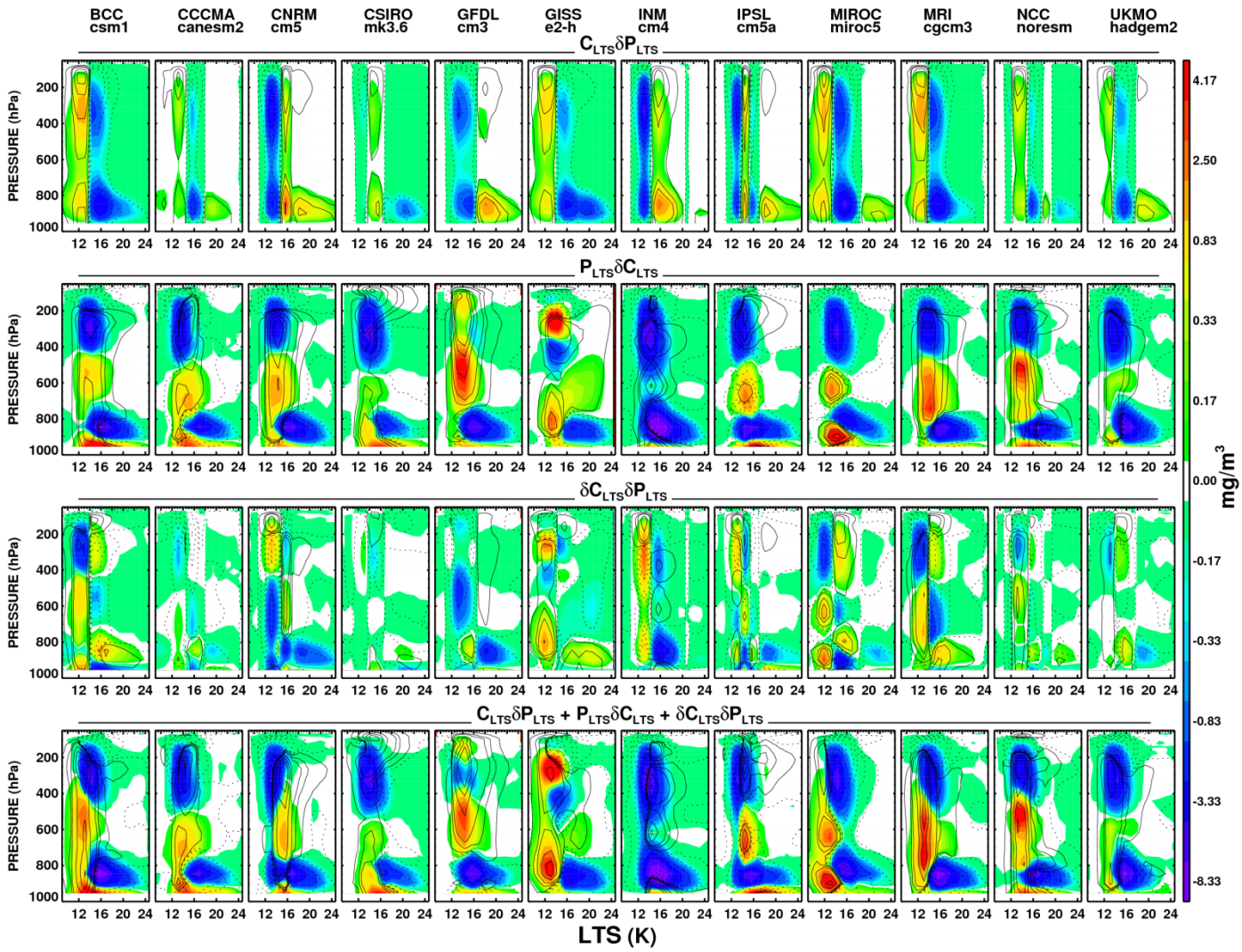
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_{LTS}\delta P_{LTS}$
$P_{LTS}\delta C_{LTS}$
$\delta C_{LTS}\delta P_{LTS}$
$C_{LTS}\delta P_{LTS} + P_{LTS}\delta C_{LTS} + \delta C_{LTS}\delta P_{LTS}$
PRESSURE (hPa)
200
400
600
800
1000
12 16 20 24
LTS (K)
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3

Figure 7b

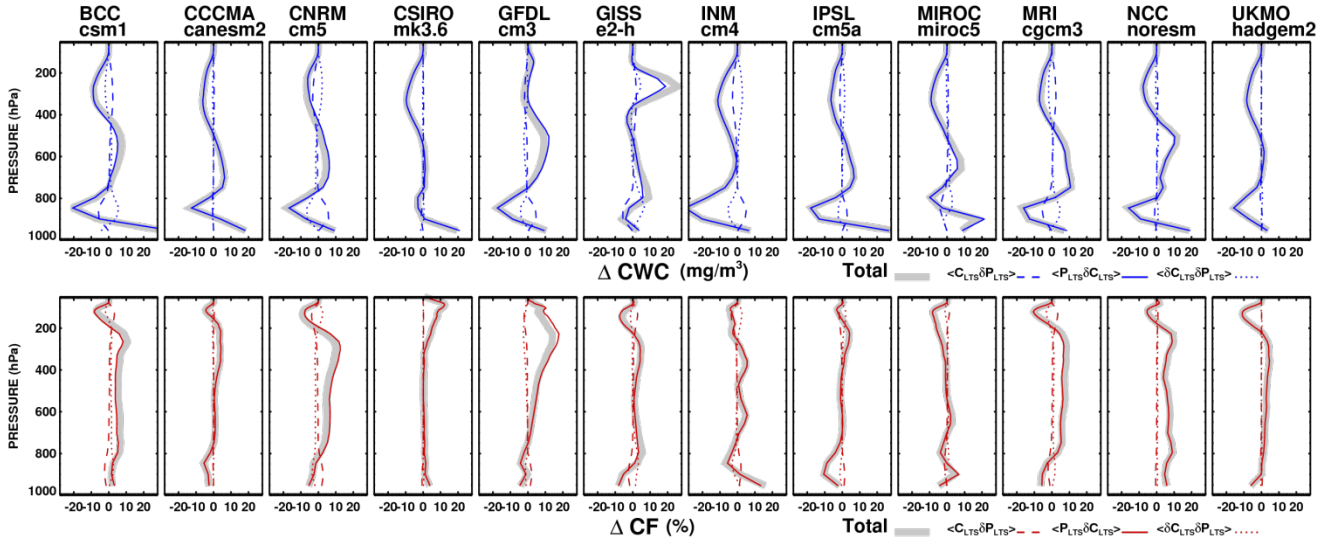
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<CLTSδPLTS>
<PLTSδCLTS>
<δCLTSδPLTS>
Δ CF (%)

Figure 7c

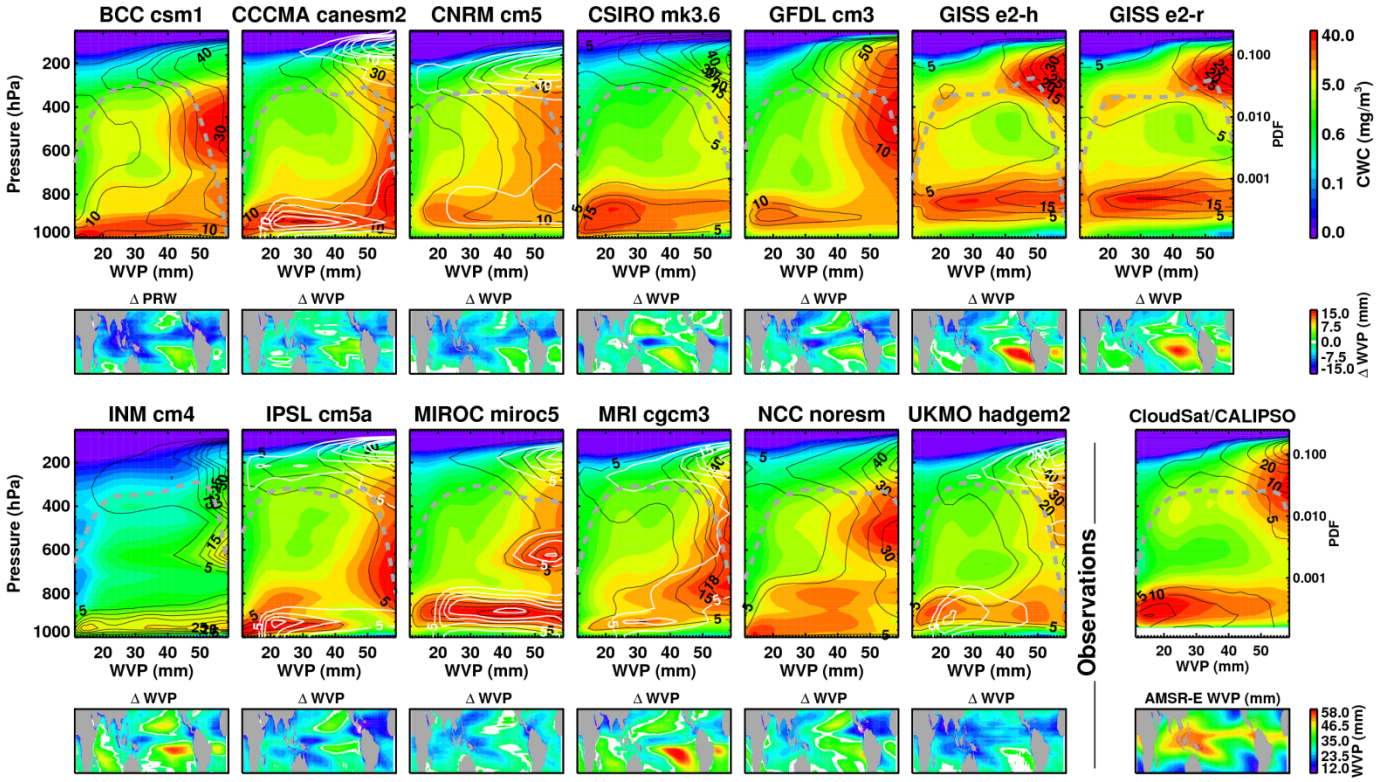
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
Observations
Pressure (hPa)
WVP (mm)
Δ PRW
Δ WVP
PDF
CWC (mg/m³)
Δ WVP (mm)
AMSR-E WVP (mm)

Figure 8a

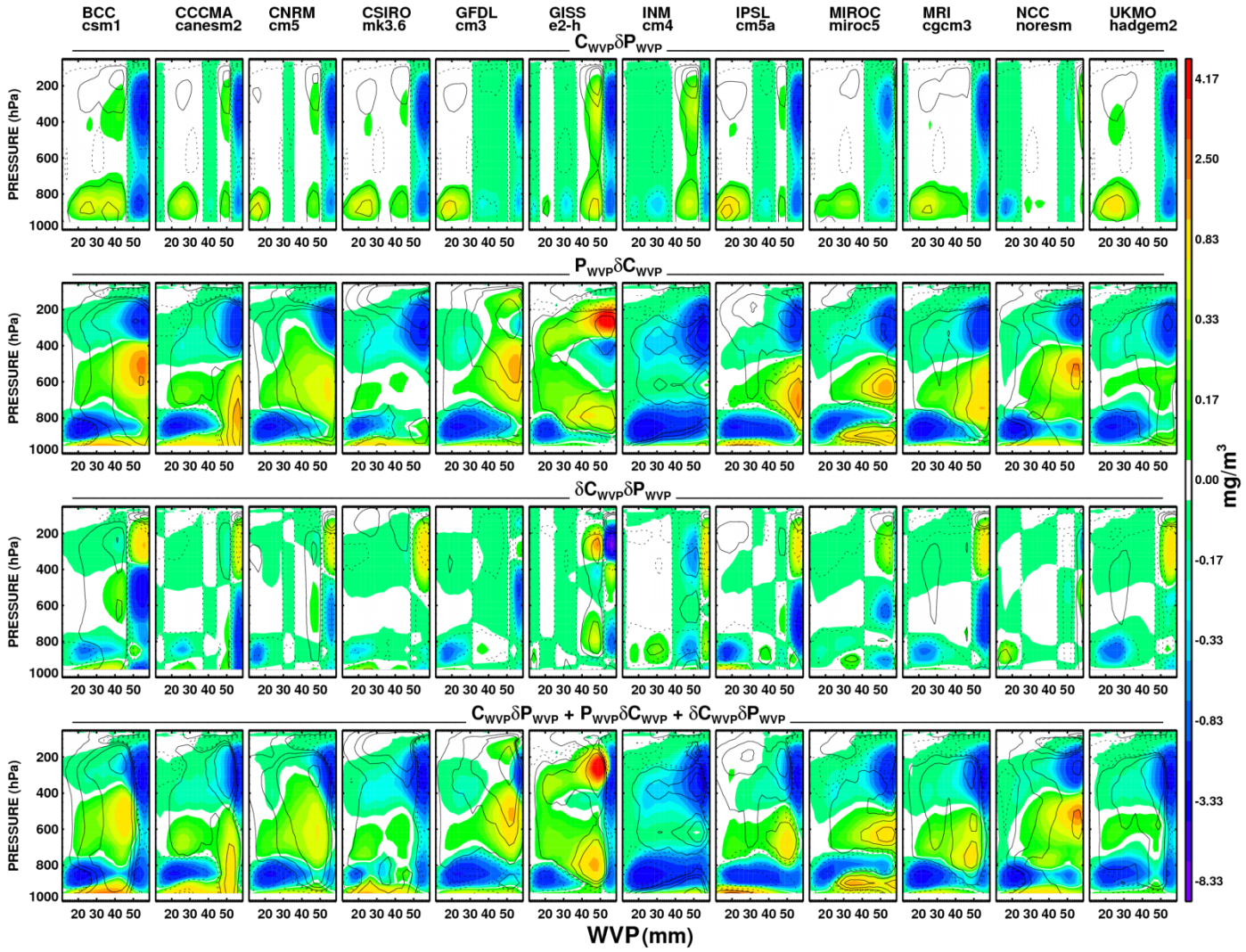
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_{WVP}\delta P_{WVP}$
$P_{WVP}\delta C_{WVP}$
$\delta C_{WVP}\delta P_{WVP}$
$C_{WVP}\delta P_{WVP} + P_{WVP}\delta C_{WVP} + \delta C_{WVP}\delta P_{WVP}$
PRESSURE (hPa)
200
400
600
800
1000
20 30 40 50
WVP (mm)
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3

Figure 8b

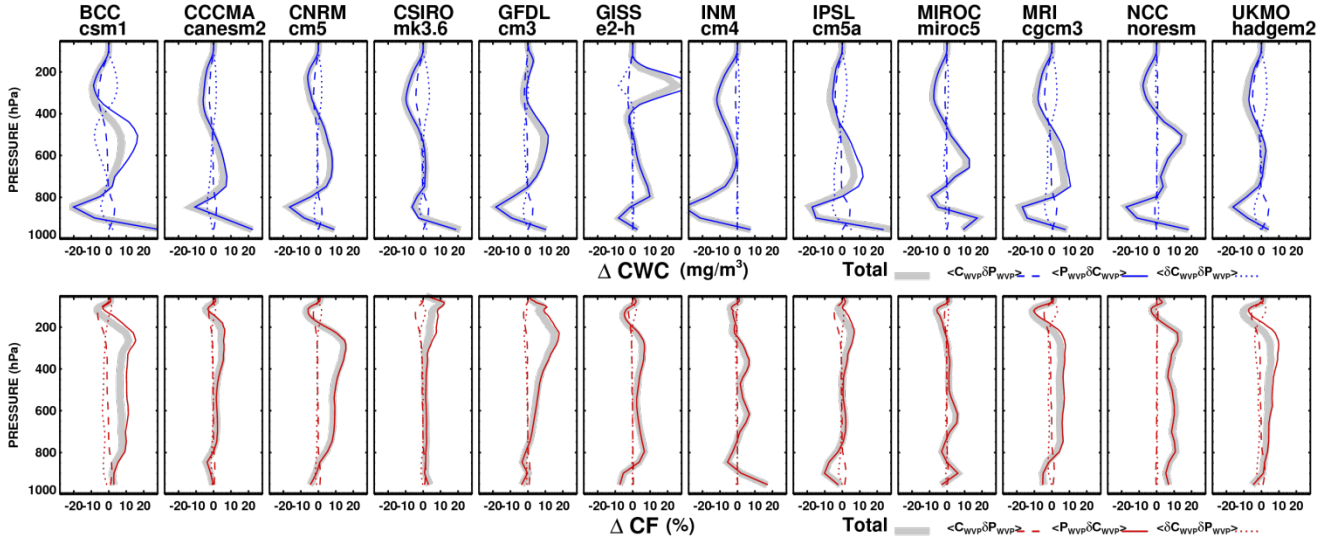
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<CwvpδPwvp>
<PwvpδCwvp>
<δCwvpδPwvp>
Δ CF (%)

Figure 8c

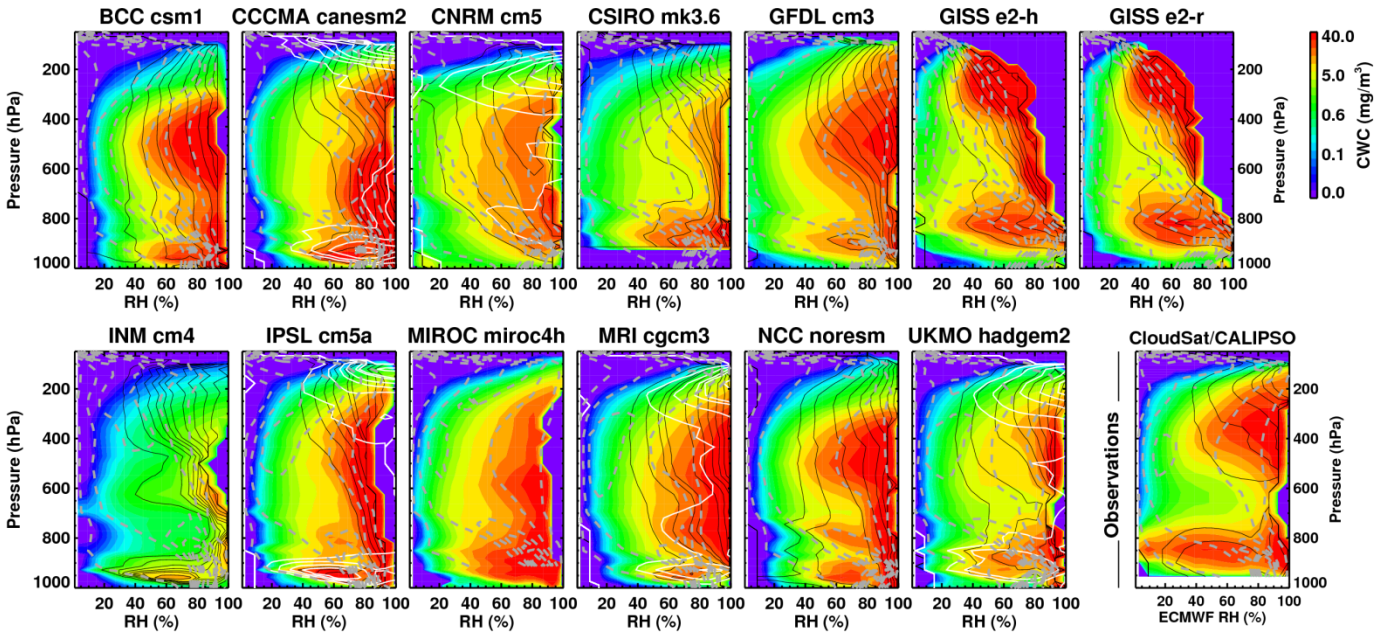
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc4h
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
Observations
Pressure (hPa)
RH (%)
ECMWF RH (%)
CWC (mg/m^3)
40.0
5.0
0.6
0.1
0.0

Figure 9a

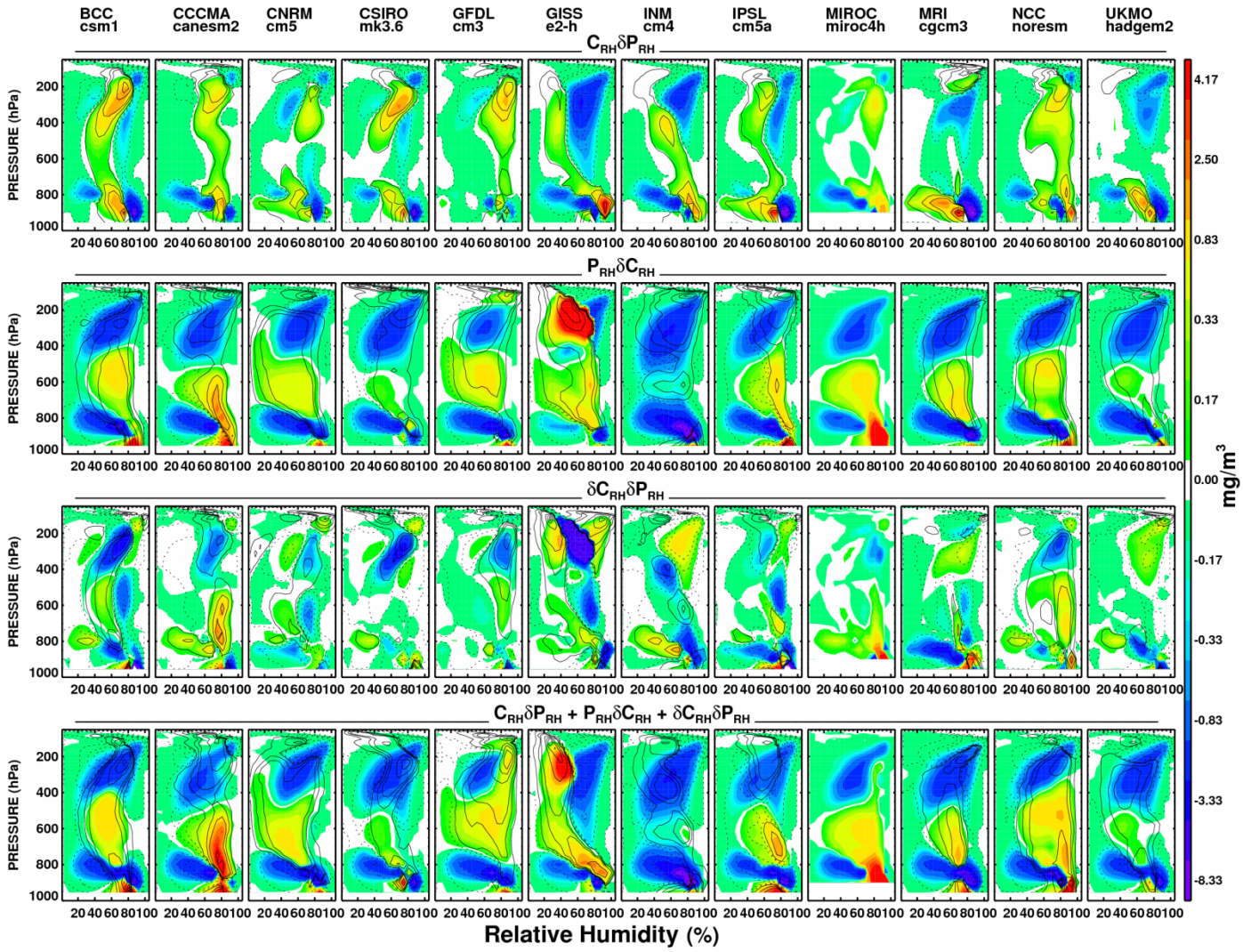
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc4h
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_{RH}\delta P_{RH}$
$P_{RH}\delta C_{RH}$
$\delta C_{RH}\delta P_{RH}$
$C_{RH}\delta P_{RH} + P_{RH}\delta C_{RH} + \delta C_{RH}\delta P_{RH}$
PRESSURE (hPa)
200
400
600
800
1000
20 40 60 80 100
Relative Humidity (%)
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3

Figure 9b

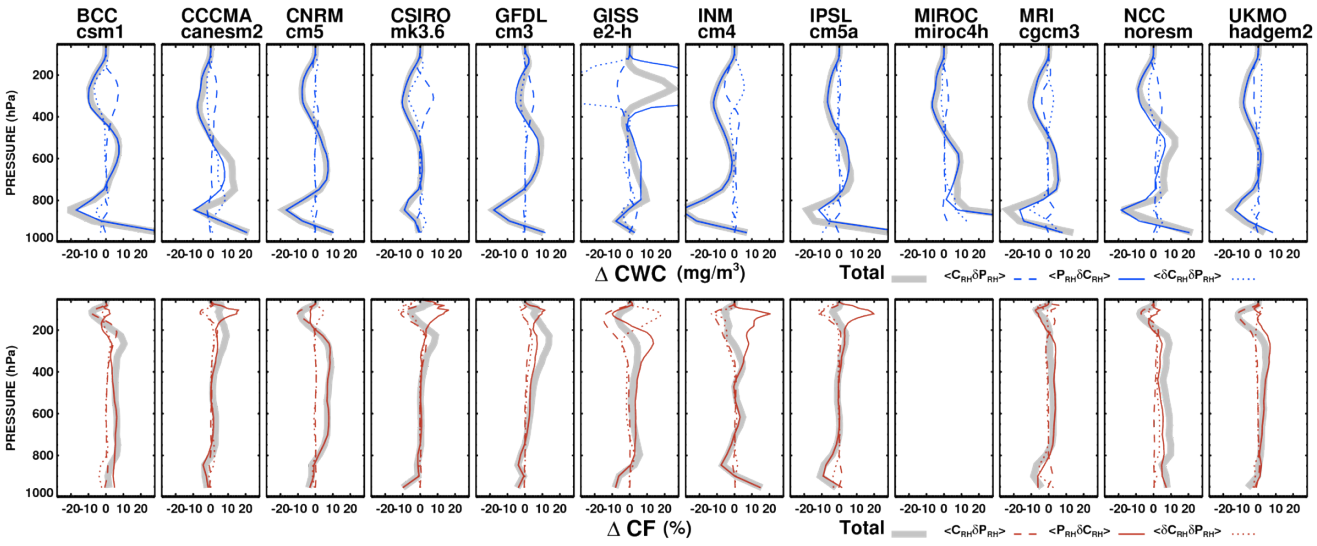
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc4h
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<CRHδPRH>
<PRHδCRH>
<δCRHδPRH>
Δ CF (%)

Figure 9c

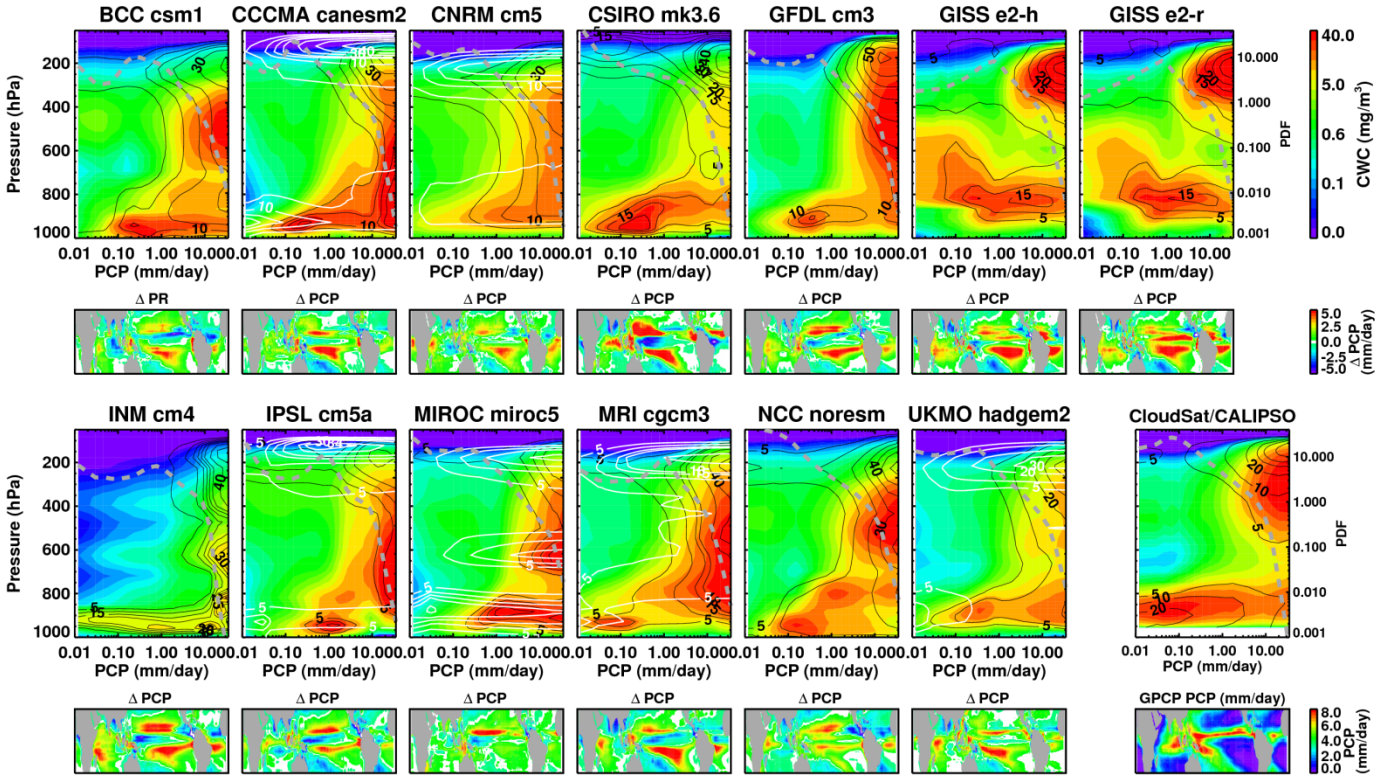
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
GISS e2-r
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
CloudSat/CALIPSO
Pressure (hPa)
PCP (mm/day)
PDF
CWC (mg/m3)
Δ PR
Δ PCP
GPCP PCP (mm/day)

Figure 10a

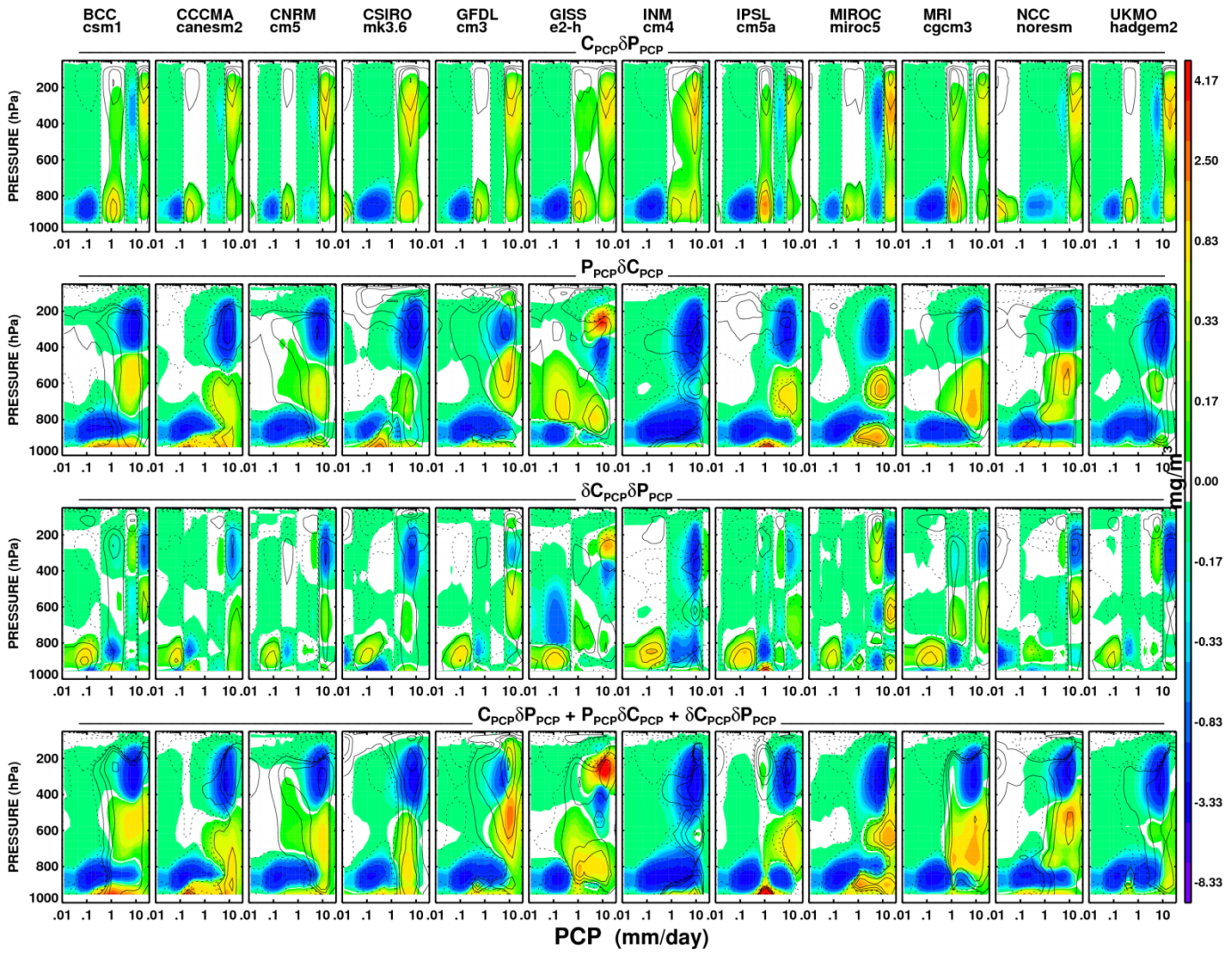
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
$C_{PCP}\delta P_{PCP}$
$P_{PCP}\delta C_{PCP}$
$\delta C_{PCP}\delta P_{PCP}$
$C_{PCP}\delta P_{PCP} + P_{PCP}\delta C_{PCP} + \delta C_{PCP}\delta P_{PCP}$
PRESSURE (hPa)
200
400
600
800
1000
.01 .1 1 10
4.17
2.50
0.83
0.33
0.17
0.00
-0.17
-0.33
-0.83
-3.33
-8.33
mg/m^3
PCP (mm/day)

Figure 10b

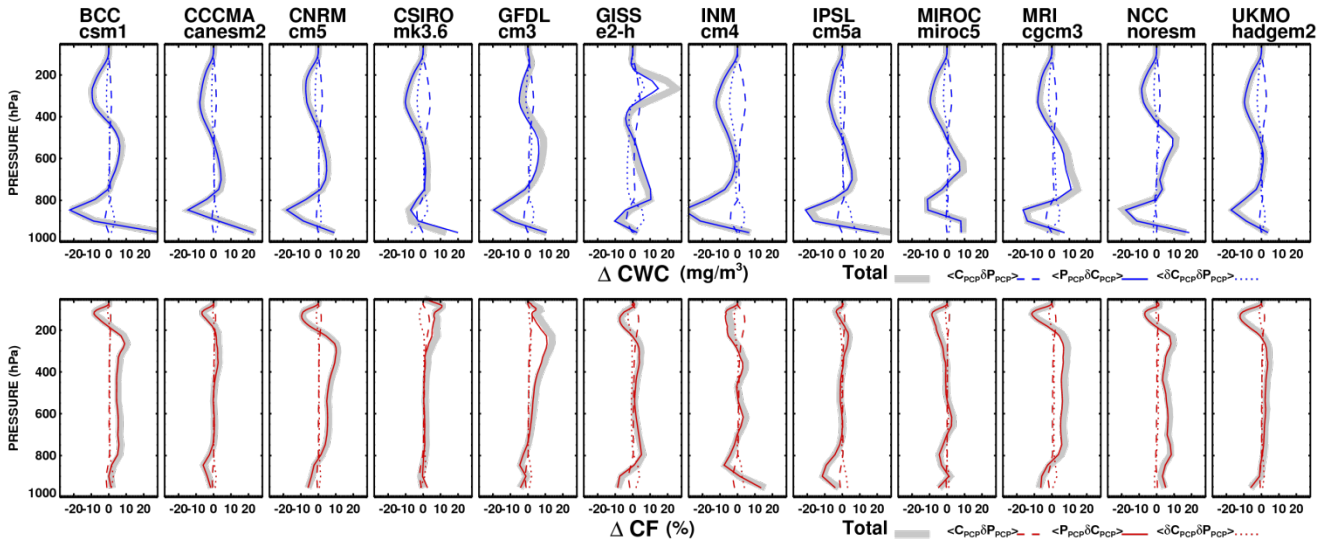
BCC csm1
CCCMA canesm2
CNRM cm5
CSIRO mk3.6
GFDL cm3
GISS e2-h
INM cm4
IPSL cm5a
MIROC miroc5
MRI cgcm3
NCC noresm
UKMO hadgem2
PRESSURE (hPa)
200
400
600
800
1000
-20 -10 0 10 20
Δ CWC (mg/m3)
Total
<CPCPδPPCP>
<PPCPδCPCP>
<δCPCPδPPCP>
Δ CF (%)

Figure 10c